DOCUMENT RESUME

ED 199 263 TM 810 157

TITLE	Procedural Handbook: 1978-79 Writing Assessment.
INSTITUTION	Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
SPONS AGENCY	National Inst. of Education (ED), Washington, D.C.
REPORT NO	NAEP-10-W-40
PUB DATE	Feb 81
CONTRACT	OEC-0-74-0506
GRANT	NIE-G-80-0003
NOTE	95p.
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Achievement Tests: Data Analysis: Data Collection: *Educational Assessment: Elementary Secondary Education: *Evaluation Methods: *National Programs: Sampling: Scoring: Test Construction: *Writing (Composition)
IDENTIFIERS	*National Assessment of Educational Progress: *Writing Evaluation

ABSTRACT

The third (1978-79) of three writing assessments conducted by the National Assessment of Educational Progress (NAEP) is reported. The writing achievement of American 9-, 13-, and 17-year-olds was surveyed using a deeply stratified, multistage probability sample design. The specific procedures used in the assessment to develop objectives and exercises, draw the assessment sample, prepare materials for the assessment, administer and score the items and analyze the results are described. In addition to reporting national results, data are provided about the performance of population subgroups within the national population, defined by sex, race, region of the country, size and type of community lived in, level of parental education, grade in school, etc. Approximately 22,500 9-year-olds, 30,500 13-year-olds, and 27,500 17-year-olds participated with between 2,100 and 2,700 students responding to each booklet. Most of the writing exercises were open-ended and of the following types: (1) holistic: (2) primary trait: (3) essay cohesion: (4) paragraph coherence: (5) syntax: and (6) mechanics. The report includes appendices, a glossary of NAEP terms and a bibliography. (RL)

PROCEDURAL HANDBOOK

1978-79 WRITING ASSESSMENT

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
Education Commission of the States

Bob Graham, Governor of Florida, Chairperson, Education Commission of the States
Robert C. Andringa, Executive Director, Education Commission of the States
Roy H. Forbes, Director, National Assessment

All National Assessment reports and publications are available through NAEP offices at the address shown at the bottom. Some of the more recent results reports are also available at the Superintendent of Documents (SOD), usually at lower prices. To order from the SOD, write to Supt. of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Check must accompany order. Allow four to eight weeks for delivery.

Reports ordered from National Assessment should be delivered within 12 days. Reports related to this report and available from National Assessment include:

WRITING

Five reports from the first assessment of writing in 1969–70
(Write to the address below for titles and prices)

2nd Assessment (1973–74)

05-W-01	*Writing Mechanics, 1969–74: A Capsule Description of Changes in Writing Mechanics,* October 1975	$ 1.60
05-W-02	*Expressive Writing,* November 1976	1.65
05-W-03	*Explanatory and Persuasive Letter Writing,* February 1977	1.85
05-W-04	*Write/Rewrite: An Assessment of Revision Skills,* July 1977	1.25
05-W-20	*The Second National Assessment of Writing: New and Reassessed Exercises With Technical Information and Data,* May 1978	25.00

3rd Assessment (1978–79)

10-W-01	*Writing Achievement, 1969–79: Results From the Third National Writing Assessment, Volume I – 17-Year-Olds,* December 1980	6.80
10-W-02	*Writing Achievement, 1969–79: Results From the Third National Writing Assessment, Volume II – 13-Year-Olds,* December 1980	
10-W-03	*Writing Achievement, 1969–79: Results From the Third National Writing Assessment, Volume III – 9-Year-Olds,* December 1980	
10-W-25	*The Third Assessment of Writing: 1978–79 Released Exercise Set,* February 1981	33.00
10-W-40	*Procedural Handbook: 1978–79 Writing Assessment,* February 1981	7.40

READING

Twelve reports from the first assessment of reading in 1970–71
(Write to the address below for titles and prices)

2nd Assessment (1974–75)

06-R-01	*Reading in America: A Perspective on Two Assessments* October 1976	1.25
06-R-21	*Reading Change, 1970–75: Summary Volume,* April 1978	2.50

ADULTS (special probe)

1st Assessment (1976–77)

08-YA-25	*Technical Information and Data From the 1977 Young Adult Assessment of Health, Energy and Reading,* March 1979	15.00
08-R-51	*Adult Readers: Will They Need Basics Too?* October 1979	1.00

LITERATURE

Six reports from the first assessment of literature in 1970-71
(Write to the address below for titles and prices)

BACKGROUND REPORT

BR-2	*Hispanic Student Achievement in Five Learning Areas: 1971–75.* Data for 9-, 13- and 17-year-olds in reading, mathematics, science, social studies and career and occupational development, May 1977	4.45

(Continued on Inside Back Cover)

PROCEDURAL HANDBOOK
1978-79 WRITING ASSESSMENT

Report No. 10-W-40

by the
National Assessment of Educational Progress

Education Commission of the States
Suite 700, 1860 Lincoln Street
Denver, Colorado 80295

February 1981

The National Assessment of Educational Progress is funded by the National Institute of Education. It is under contract with the Education Commission of the States. It is the policy of the Education Commission of the States to take affirmative action to prevent discrimination in its policies, programs and employment practices.

Library of Congress
Catalog Card Number: 72-183665

Although a few early National Assessment reports have individual catalog card numbers, all recent reports have been assigned the above series number.

ISBN 0-89398-397-7

The National Assessment of Educational Progress is an education research project mandated by Congress to collect and report data, over time, on the performance of young Americans in various learning areas. National Assessment makes available information on assessment procedures and materials to state and local education agencies and others.

The work upon which this publication is based was performed pursuant to Contract No. OEC-0-74-0506 of the National Center for Education Statistics and the National Institute of Education; also, Grant No. NIE-G-80-0003 of the National Institute of Education. It does not, however, necessarily reflect the views of those agencies.

TABLE OF CONTENTS

LIST OF TABLES

FOREWORD

When the U.S. Office of Education was chartered in 1867, one charge to its commissioners was to determine the nation's progress in education. The National Assessment of Educational Progress (NAEP) was initiated a century later to address, in a systematic way, that charge.

Since 1969, the National Assessment has gathered information about levels of educational achievement across the country and reported its findings to the nation. It has surveyed the attainments of 9-year-olds, 13-year-olds, 17-year-olds and adults in art, career and occupational development, citizenship/social studies, mathematics, music, reading/literature, science and writing. All areas have been periodically reassessed in order to detect any important changes. To date, National Assessment has interviewed and tested more than 900,000 young Americans.

Learning-area assessments evolve from a consensus process. Each assessment is the product of several years of work by a great many educators, scholars and lay persons from all over the nation. Initially, these people design objectives for each subject area, proposing general goals they feel Americans should be achieving in the course of their education. After careful reviews, these objectives are given to exercise (item) writers, whose task it is to create measurement instruments appropriate to the objectives.

When the exercises have passed extensive reviews by subject-matter specialists, measurement experts and lay persons, they are administered to probability samples. The people who compose these samples are chosen in such a way that the results of their assessment can be generalized to an entire national population. That is, on the basis of the performance of about 2,500 9-year-olds on a given exercise, we can make generalizations about the probable performance of all 9-year-olds in the nation.

After assessment data have been collected, scored and analyzed, the National Assessment publishes reports and disseminates the results as widely as possible. Not all exercises are released for publication. Because NAEP will readminister some of the same exercises in the future to determine whether the performance level of Americans has increased, remained stable or decreased, it is essential that they not be released in order to preserve the integrity of the study.

ACKNOWLEDGMENTS

Many organizations and individuals have made substantial contributions to the writing assessments. Not the least of those to be gratefully acknowledged are the administrators, teachers and students who cooperated so generously during the collection of the data.

Special acknowledgment must go to the many writing educators and specialists who provided their expertise in the development, review and selection of the assessment objectives and exercises. Particular thanks are given to Carl Klaus and Richard Lloyd-Jones, both of the University of Iowa, for providing leadership in the development of primary trait exercises and scoring guides; to John Mellon, University of Illinois at Chicago Circle, for providing leadership in the development of the sentence combining exercises and the mechanics and syntax scoring guides; to Charles Cooper, University of California at San Diego, for providing leadership in the development of the cohesion scoring guides; and to Paul Diehl, University of Iowa, for providing special assistance in all areas.

Administration of the writing assessments was conducted by the Research Triangle Institute, Raleigh, North Carolina. Scoring and processing were carried out by the Westinghouse DataScore Systems (formerly Measurement Research Center), Iowa City, Iowa, and by the National Assessment staff. Sue Worthen, Donna Benson and Dan Duse deserve special mention for their excellent work supervising the primary trait, cohesion, syntax and mechanics scoring, as does Wendy Littlefair for her consulting work with the syntax and mechanics scoring. Edward White, University of California at San Bernardino, and his holistic scoring staff also deserve thanks.

Every assessment is the result of a collaborative effort by the National Assessment staff. Particular thanks are extended to Ina Mullis of the Analysis and Research Department who had major responsibility for coordinating the development, scoring and analysis of the writing assessment. Special thanks are also given to Rex Brown, Director of the Publications and User Products Department, for his contributions to exercise development and reporting and to Jim Damon for data processing support. Acknowledgment is also due to the following people: Donald T. Searls and Eugene Johnson for information on sampling and data analysis, Dunlap Scott for information on data collection; Ava Powell for technical support; Marci Reser and Carmen Nietes for production of this report, and Barbara Ward for editorial supervision.

Roy H. Forbes
Director

INTRODUCTION

The National Assessment of Educational Progress (NAEP) has completed three assessments of writing, the first conducted in 1969-70, the second in 1973-74 and the third during 1978-79. Each assessment surveyed the writing achievement of American 9-, 13- and 17-year-olds, using a deeply stratified, multistage probability sample design. This report documents procedures used in the 1978-79 writing assessment and also describes changes in procedures between assessments.

To measure the status of writing achievement in 1978-79, National Assessment consultants first reviewed and refined the objectives used in the second writing assessment. Consultants then worked with staff to develop exercises. Most items (exercises)[1] developed for the 1978-79 assessment will be kept secure and used for measuring changes in performance in future assessments, but some will be released following the 1978-79 data collection.

Approximately 22,500 9-year-olds, 30,500 13-year-olds and 27,500 17-year-olds[2] participated in the 1978-79 writing assessment. Because each item booklet is administered to a representative sample at an age group and because data are reported only for groups of students, not individuals, each student assessed did not respond to all items. Each student completed one item booklet of about 45 minutes in length. Between 2,100 and 2,700 students responded to each booklet.

The exercises for each assessment were administered by a professional data collection staff to minimize the burden on participating schools and to maximize uniformity of assessment conditions. Instructions and items were recorded on a paced audio tape and played back to students to reduce the potential effect of reading difficulties and to insure that all students moved through the booklets at the same speed.

Since National Assessment staff and consultants feel strongly that writing performance should be assessed on the basis of writing samples rather than objective tests, most of the writing exercises are open-ended, requiring students to produce a piece of writing. Responses to open-ended items were

---

[1]National Assessment uses the term "exercise" to mean an assessment item. The terms "exercise" and "item" are used interchangeably in this report.

[2]For the 1978-79 assessment, 17-year-olds sampled included only those attending school at the time of the assessment.

hand-scored by trained readers using detailed scoring guides; the few multiple-choice items included were scored by an optical scanning machine. Several types of open-ended scoring were conducted -- holistic, primary trait, essay cohesion, paragraph coherence, syntax and mechanics. For each type of scoring procedure, papers collected in different assessments were randomly ordered into a single pool and scored simultaneously following the 1978-79 assessment. Different types of scoring were done by different groups of scorers.

When a report indicates that "85% of the 17-year-olds gave a particular response," it means that an estimated 85% of the 17-year-olds would have given that response if all the 17-year-olds in schools across the country had been assessed. National Assessment reports of writing results provide estimates of percentages of responses at each score level or some combination of score levels for each writing task. Increases or decreases in these percentages between assessments are used to indicate trends in achievement. To provide descriptive information about syntax and mechanics, data are presented that illustrate the average performance as well as the range of performance. In addition to reporting national results, National Assessment provides data on the performance of various population subgroups within the national population, defined by sex, race, region of the country, size and type of community lived in, level of parental education, grade in school and other categories as appropriate.

The following chapters describe specific procedures used to develop objectives and exercises, draw the assessment sample, prepare materials for the assessment, administer and score the items and analyze the results.

CHAPTER 1

OBJECTIVES REDEVELOPMENT

The primary goal of the National Assessment of Educational Progress (NAEP) is to report on the current educational status of young Americans and to monitor any changes in achievement over time. For each learning area to be assessed, NAEP asks consultants to develop objectives that define the subject area. Since the objectives provide guidelines for exercise developers, consultants are asked to include examples of the knowledge, skills and attitudes to be assessed at each age level.

Education in America is a collaborative enterprise involving a great many people with widely differing philosophies. Providing information about education nationwide would be considerably easier if there were consensus about the means and ends of American education, but the fact is that Americans have conflicting and sometimes contradictory values regarding the goals of education and the means for achieving them. To develop an assessment that is truly national in scope and takes into account the diversity of curricula, values and goals across the country, National Assessment employs a consensus process for developing objectives, with representation of many different groups of people.

Several types of consultants help to develop National Assessment objectives. College and university specialists in a learning area insure that the objectives include important concepts that the schools should be teaching. Educators, including classroom teachers, curriculum supervisors and persons involved in teacher education, make sure that the objectives describe concepts, skills and attitudes that the schools should be teaching and those that they presently are teaching. Concerned citizens, parents and other interested lay persons must agree that the objectives are important for young people to achieve, are free of educational jargon and are not biased or offensive to any groups. Consultants are selected to represent different regions of the country, minority groups, various types of communities and age levels.

Table 1 presents the objectives used for the second and third writing assessments. The objectives for the 1973-74 assessment also served as the basis for the 1978-79 writing assessment, since restricted funds and limited development time made a full-scale revision effort impractical. However, the objectives were refined based on NAEP's experiences with the second writing assessment and comments from users of NAEP materials.

Objectives II and III are based on the writing objectives developed for

the 1969-70 assessment by the Educational Testing Service (ETS), Princeton, New Jersey, in 1965. The original writing objectives were reviewed by writing experts and lay panels and used as the basis for generating exercises for the first assessment of writing.

TABLE 1. Objectives for the 1973-74 and 1978-79 Writing Assessments

I. Demonstrates ability in writing to reveal personal feelings and ideas

   A. Through free expression
   B. Through the use of conventional modes of discourse

II. Demonstrates ability to write in response to a wide range of societal demands and obligations. Ability is defined to include correctness in usage, punctuation, spelling and form or convention as appropriate to particular writing tasks, e.g., manuscripts, letters

   A. Social
      1. Personal
      2. Organizational
      3. Community
   B. Business/vocational
   C. Scholastic

III. Indicates the importance attached to writing skills

   A. Recognizes the necessity of writing for a variety of needs (as in Objectives I and II)
   B. Writes to fulfill those needs
   C. Gets satisfaction, even enjoyment, from having written something well

---

Redevelopment of objectives for the second (1973-74) assessment of writing was also done by ETS. Recommendations from consultants indicated that creative or personal writing as well as expository writing should be a part of the assessment. Thus, Objective I was added. Also, lay reviewers felt that the mechanical aspects of writing, such as grammar, punctuation and spelling, should be included in the revised description of Objective II, although they agreed that mechanical correctness was not to be regarded as the only criterion for judging a piece of writing.

A recurring concern has been the difficulty and appropriateness of measuring an objective that requests respondents to reveal personal feelings. One suggestion was that Objective I might apply to personal writing done with feeling; another was that the objective might simply require the expression of feelings (whether they were "fictional" or "real"). National Assessment has elected to interpret Objective I broadly as the ability to engage in writing for expressive purposes.

Prior to the development of second assessment exercises, National Assessment staff and consultants decided that writing, particularly public writing of the sort described in Objective II, is done for a particular audience and should be evaluated in view of its intended effect on that audience. This decision necessitated the development of narrowly defined tasks that required writing to specified audiences for particular purposes. The major purpose of most social, business and scholastic writing was identified as either persuasive or explanatory. Thus, for the second and third assessments of writing, National Assessment developed Objective II writing tasks to measure either explanatory or persuasive writing appropriate to specific social, business or scholastic situations.

As a result of current assessment procedures and available resources, the last two assessments have concentrated on Objectives I and II. This is not meant to imply that Objective III is unimportant.

For a complete description of the writing objectives and procedures used to develop them, see Writing Objectives, Second Assessment (1972).

CHAPTER 2

DEVELOPMENT OF EXERCISES

The Exercise Development Framework

National Assessment's major task is to describe changes over time in the educational attainments of students. To do this effectively, NAEP must collect information about what things students know and can do in a variety of subjects. To provide such information about writing skills, National Assessment found it necessary to develop a system for assessing writing that would provide data about specific skills within a reasonable budget. For the first assessment of writing, essay tasks were scored holistically. This approach to scoring involves the use of several broad scoring categories with papers ranked according to the general quality of an essay taken as a whole. However, this approach did not provide sufficient descriptive information about specific writing skills for National Assessment's purposes.

Prior to the second assessment of writing, NAEP staff held a conference of writing educators and measurement specialists. These specialists indicated that both evaluative and descriptive information about writing were needed. Since most writing is done with a purpose in mind, they proposed that the assessment concentrate on measuring the writing abilities needed to perform particular tasks. Thus, beginning with the 1973-74 writing assessment, NAEP has used the Primary Trait System (PTS) for scoring essays, which evaluates the capacity to write for precisely defined purposes. In this system, an overall evaluative rating on a 4-point scale indicates level of achievement of the primary trait defined for the writing task -- that is, how successful each paper is in reaching its purpose. Scoring for secondary characteristics, such as control of tense, coherence, mechanics or consistency of voice, provides descriptive data.

Such an evaluation system has implications for exercise and scoring guide development. First, to insure that the writing tasks require a range of skills, exercise developers need guidelines about the kinds of tasks that should be developed. Accordingly, National Assessment staff held conferences with writing educators to refine the objectives and provide new guidelines for writing task development. Objective I was interpreted as the ability to engage in writing for expressive purposes, while Objective II was interpreted as requiring persuasive or explanatory writing. It was agreed that each age group should be asked to write at least one expressive, one persuasive and one explanatory task, and that the tasks would be varied over the years. Tasks developed to measure expressive writing have included narratives,

role-playing, fantasy, expression of emotion and humorous writing. Persuasive and explanatory tasks have required letters, speeches and reports appropriate to various social, business and scholastic situations -- for example, a "thank you" letter, an accident report, a speech to a community or a letter correcting a billing error.

Second, each exercise must clearly explain the writing task required. It must be specific as to the role of the writer, the purpose of the communication and the intended audience, and the evaluative criteria to be used by scorers should be made apparent to the writer. For example, if papers are to be judged on expressive qualities, writers should be instructed to try to be expressive. Furthermore, questions and rhetorical situations should be phrased in such a way that all respondents will give comparable answers.

Third, to insure that responses can be evaluated in a way that provides information about the skill the exercise is attempting to measure, the rating criteria must be developed along with the exercise. The primary trait score levels and, when possible, procedures for rating designated secondary characteristics, should be identified in advance of field testing and then refined using sample papers.

1978-79 Exercise Development

Exercises were developed by persons active in the field of writing education. The main developmental effort focused on primary trait system exercises designed to assess the skills necessary for effective expressive, explanatory and persuasive writing. Since National Assessment does not have the resources to assess a wide variety of tasks at any one time, there was an effort to cover some areas not emphasized in earlier assessments. There was also an effort to include some items that examined use of commonly accepted strategies for improving writing. For instance, one item asked 13- and 17-year-olds to revise their work after writing a draft; another asked them to "prewrite" or jot down ideas before beginning to write.

In designing exercises for the 1978-79 writing assessment, developers also investigated the possibility of using more multiple-choice items to assess writing skills. Some items assessing ability to locate the central idea of a paragraph and ability to identify words needed to make up correct grammatical structures in sentences were developed but not assessed because consultants indicated that this type of information could be obtained from standardized test data. Consultants felt that assessment resources would be better devoted to obtaining information about these skills from actual writing samples; thus, two new scoring systems, one for cohesion and one for syntax, were developed (see Chapter 6).

Two types of items other than essay tasks developed for the 1978-79 assessment did prove successful: sentence combining tasks and paragraphs with connective ties clozed. These exercises were designed not to measure specific objectives but rather to provide additional information about particular writing skills and to help in providing a context for reporting changes in writing performance.

The sentence combining tasks required respondents to combine two or more simple sentences into a longer sentence that meant the same thing. In the cloze paragraph items, a paragraph was constructed to contain a number of pronoun references and specific transitional devices. Several of these references and connective ties were clozed (deleted and replaced by blanks), and respondents were offered a number of multiple-choice options to replace each blank.

Finally, to find out about students' attitudes toward writing and to provide some instructional context for reporting results, an attitudinal scale and additional background questions were developed.

Released Writing Tasks

About half of the 1978-79 writing assessment at each age consisted of reassessed items already given in previous assessments so that measurements of changes in performance could be made. Several of these items appeared in all three assessments; others were administered only in the second and third assessments.

The last of the items developed for the first assessment and most of the items assessed in both 1973-74 and 1978-79 have been released and published in Report 10-W-25, The Third Assessment of Writing: 1978-79 Released Exercise Set (1981). Some items developed for the 1973-74 assessment and most of the new items developed for the 1978-79 assessment will be kept secure and used to measure changes in achievement in future assessments. Table 2 shows essay and letter writing items released after the third assessment and the years in which they were assessed. Each task is labeled with a short name and unique item identification number. ID numbers beginning with "1" indicate the item measures expressive writing; numbers beginning with "2" indicate items measuring explanatory or persuasive writing.

Sentence combining and cloze paragraph items were assessed for the first time in 1978-79; about half of these have been released and the remainder will be used to measure change.

Item Development, Field Test, Review and Selection Procedures

Item writers worked in small groups to create items, and the groups then critiqued each other's work. Once items had been developed, critiqued and revised, they were reviewed by National Assessment staff. The results of these reviews were compiled and item writers once again revised items. Surviving items were field tested in schools across the country to discover potential problems in wording, directions or administrative procedures and to collect item statistics, timing information and scoring information. Schools were selected to represent high- and low-income communities as well as more typical communities. The tryouts were administered to students in at least four classrooms (approximately 100 students) at each of the ages assessed.

TABLE 2. Essay and Letter Writing Exercises
Released Following 1978-79 Assessment

Assessed in 1969-70, 1973-74 and 1978-79		
	Age 9	Fireflies (102012)
		Kangaroo (102013)
	Age 13	Rainy Day (102015)
	Age 17	Stork (102016)
	Ages 13 and 17	Describe Something (203012)
Assessed in 1973-74 and 1978-79		
	Age 9	Goldfish (101006)
		Puppy Letter (201001)
	Age 13	Loss (101007)
		Principal Letter (201006)
	Age 17	Grape Peeler (101015)
		Rec Center (201007)
		Electric Blanket (202014)
Assessed in 1978-79		
	Ages 9 and 13	Poster Calendar (202031)

So that the field tests would closely simulate actual assessment field procedures, the students recorded their answers in the test booklets; directions and questions were read to students from an audio tape and National Assessment staff members, rather than classroom teachers, administered the test. The students' responses to the items, as well as the administrators' reports of any field problems, helped both staff and consultants to evaluate and revise the exercises. Revised exercises were generally field tested again.

After exercises were field tested, the results were reviewed by NAEP staff and panels of writing educators and lay persons from across the country. Exercises for each age group were reviewed for appropriateness by teachers who taught students at that age. Lay citizens, representing a variety of occupations and interests, also reviewed the exercises, checking for sex or racial bias and considering the general importance of each exercise. A panel of writing specialists worked with staff to examine the items judged to be successful by the review panels and then made the final selection of the items included in the 1978-79 assessment.

CHAPTER 3

PREPARATION OF ASSESSMENT MATERIALS

Preparation of Booklets and Audio Tapes

National Assessment uses a matrix sampling approach, with different nationally representative samples of students responding to different item booklets (see Chapter 4 for details). Since the Assessment's aim is to describe results for groups of students (males, blacks, students in the West, and so on), not individuals, it is not necessary for each student to respond to all the items. Each student responded to one booklet of items designed to be completed in a single class period.

The 1978-79 assessment was a combined assessment of writing, music and art. Because of the length of many of the writing items, booklets of exercises included either writing and music exercises or writing and art exercises. Following the selection of exercises to be included in the assessment, National Assessment staff determined which exercises were to be included in the various booklets and sequenced them within the booklets. Booklets were constructed separately for each age level since students at different ages received different sets of exercises. Thus, exercises for 9-year-olds were not sequenced in the same order as those for 13-year-olds, and so forth. In 1978-79, there were 9 exercise booklets that contained writing exercises for 9-year-olds, 11 such booklets at age 13 and 10 such booklets for 17-year-olds.

The following constraints were observed in preparing the 1978-79 exercise booklets:

1. Each booklet contained exercises of varying difficulty so that students would not become bored by many easy exercises or discouraged by many difficult exercises.

2. Exercises could not cue other exercises. In other words, the answer to one exercise could not be contained in another exercise in the same booklet.

3. Each booklet was timed so that it would take no more than 45 minutes -- the length of a typical class period -- of a student's time. Booklets contained approximately 30-35 minutes of exercise time and an additional 10-15 minutes of introductory material, instructions and background questions.

4. Booklets were designed to be, insofar as possible, parallel with respect to the number of different objectives measured and difficulty levels. Items measuring a particular objective were scattered throughout the booklets so that many different students would respond to questions related to a particular objective.

National Assessment has constantly attempted to institute procedures to minimize difficulties connected with the testing situation so that results will be, as nearly as possible, an accurate reflection of what students know and can do. Considerable effort was devoted to developing clear instructions for procedures students should use in answering items and to formating the booklets to help students write as well as possible in the assessment situation. For example, space was provided for students to write their answers directly in the assessment booklets, not on separate answer sheets. It was felt that this procedure would reduce possibilities for confusion in using additional sheets of paper, especially for the younger students. Also, to reduce the possibility of fatigue effects, essay items were placed in the beginning of booklets so students could write earlier rather than later in the administrative session. To minimize guessing, students were encouraged to select the "I don't know" response option included with multiple-choice items.

Paced audio tapes were prepared for each exercise booklet. Most of the written portions of an exercise stimulus and response options were read aloud to reduce the effect of any reading difficulties and to insure that all students moved through the booklets at the same speed. In addition, the use of tapes helped to insure uniform assessment conditions across the country.

In 1973-74 and 1978-79, special instructions were given to item administrators regarding the longer (15-25 minutes) writing tasks. If the students appeared to be finished with an item well before the allotted time had been used, administrators were asked to announce that students had several minutes to finish up and then to move on to the next item. As it turned out, administrators seldom used this procedure but felt it was extremely important to have such an option for those cases when a group did finish early. National Assessment has found that if the administrator cannot proceed, the respondents may get restless or even leave the assessment session.

Differences in Item Booklets in the Three Writing Assessments

National Assessment makes every effort to make assessment conditions for items measuring change identical from assessment to assessment so that any changes observed will be attributable to changes in achievement rather than a response to an altered testing condition. Although items were kept identical, the makeup of the item booklets was different for each writing assessment. In 1969-70, writing items were assessed with science and citizenship items; in 1973-74, writing and career and occupational development items were included in the same booklets; in 1978-79, writing items shared booklets with either art or music items. Typically only one writing essay was included in an item

booklet. Although some essays were given at the end of administrative sessions in the first assessment, writing essays were consistently placed very close to the beginning of the booklets in the second and third assessments. This change was made to reduce the effects of differences in context from assessment to assessment.

CHAPTER 4

SAMPLING

This chapter describes procedures used in designing and selecting the samples for the three writing assessments. Sample design and selection for the writing assessments were conducted by the Research Triangle Institute, Raleigh, North Carolina, and monitored by NAEP staff.

The target populations for each of the assessments included 9-, 13- and 17-year-olds[1] enrolled in either public or private schools at the time of the assessment who were not functionally handicapped to the extent that they could not participate in an assessment. Specific groups excluded were: non-English-speaking persons, those identified as nonreaders, persons physically or mentally unable to respond, and persons in institutions or attending schools established for the physically or mentally handicapped.

National Assessment did not follow up specific individuals from one assessment to the next. In other words, the students who participated in the first writing assessment were not the same ones who participated in the second or third assessment. However, in each assessment year, participants were carefully selected to represent each age level. For example, at age 9 different sets of probability samples were used for the three assessments, but each set contained nationally representative samples of the population of students who were nine years old during that assessment year. Thus, if we say that 9-year-olds' achievement improved between 1974 and 1979, we mean that students who were nine years old in 1979 correctly answered the same questions more often than those who were nine years old in 1974.

The NAEP samples were designed to provide approximately 2,000 to 2,700 respondents per writing task. These numbers allowed reporting of data for the nation and for the subgroups defined in Appendix A. The definitions of the target populations were identical in each assessment; however, the design used to obtain representative samples of these populations was modified following the second writing assessment.

Several essays were scored for mechanical errors and use of various syntactic devices in addition to primary trait or cohesive scoring. While NAEP

---

[1]Definition of 1978-79 assessment age groups are: 9-year-olds -- born during calendar year 1969; 13-year-olds -- born during calendar year 1965; and 17-year-olds -- born October 1, 1961, through September 30, 1962.

feels that descriptions of mechanics and syntax provide valuable information, such detailed scoring is also quite expensive. Accordingly, only subsamples of responses to several essays were scored in this fashion. Use of a subsample allows reporting of results for the nation but for only a limited number of subgroups.

An overview of the National Assessment approach to sample design is given below, followed by a description of procedures specific to the three writing assessments and a description of the procedures used to obtain the subsamples of papers.

Overview of the National Assessment Sample Design

For all of its assessments, National Assessment uses a deeply stratified, three-stage national probability sample design with oversampling of low-income and rural areas. In the first stage, the United States is divided into geographical units of counties or groups of contiguous counties meeting a minimum population size requirement. These units, called primary sampling units (PSUs), are stratified by region and size of community. From the list of PSUs, a sample of PSUs is drawn without replacement with probability proportional to population size measures, representing all regions and sizes of communities. Oversampling of low-income and extreme-rural areas is first performed at this stage by adjusting the estimated population-size measures of such areas to increase sampling rates. In the current sampling procedures, Census Employment Survey Data are used within PSUs to further delineate and oversample low-income areas. Counties with high proportions of rural families are also oversampled.

In the second stage, all public and private schools within each PSU selected in the first stage are listed. Schools within each PSU are selected without replacement with probabilities proportional to the number of age-eligibles in the school.

The third stage of sampling occurs during the data collection period. A list of all age-eligible students within each selected school is made. A simple random selection of eligible students without replacement is obtained, and item booklets are administered to selected students. Specially trained personnel select the sample and administer the booklets.

Each respondent in the sample does not have the same probability of selection because some subpopulations are oversampled and because adjustments are made to compensate for student nonresponse and for some schools' refusal to participate. The selection probability for each individual is computed, and its reciprocal is used to weight each response in any statistical calculation to compensate for unequal rates of sampling and to insure proper representation in the population structure.

The number of PSUs, schools within PSUs and students within schools are determined by optimum sampling principles. That is, a sample design is

utilized that will achieve the maximum precision for a given level of resources.

Table 3 displays the number of PSUs used and the number of schools in which assessment sessions were conducted by age for each of the three writing assessments.

TABLE 3. Number of PSUs Selected and Schools Within PSUs Included in the Assessment in 1969-70, 1973-74 and 1978-79

	1969-70 Assessment		1973-74 Assessment		1978-79 Assessment	
	No. of PSUs	No. of Schools	No. of PSUs	No. of Schools	No. of PSUs	No. of Schools
Age 9	204	935	116	1,246	75	539
Age 13	205	749	116	1,278	75	496
Age 17	193	670	116	1,052	75	435

Changes in Sample Design: 1969-70, 1973-74 and 1978-79

A major change in sampling procedures has been a reduction in the target populations with each subsequent assessment. In 1969-70 and 1973-74, 17-year-olds who were not currently attending school were included in the assessment. These 17-year-olds were sampled by getting their names from school lists of dropouts and early graduates. Young adults aged 26-35, identified through a household sample, were assessed in the 1969-70 assessment. Assessment of out-of-school 17-year-olds and adults is quite expensive, and in 1978-79 funds to assess these groups were not available.

Procedures for selecting the samples of students attending school have changed somewhat between the writing assessments. For the 1978-79 sample, two types of PSUs were identified: (1) large-size population areas defined by the U.S. Bureau of the Census as Standard Metropolitan Statistical Areas (SMSAs) and (2) other contiguous non-SMSA counties grouped together to meet certain minimum size requirements. The first stratification of PSUs was by geographic region, as defined by the Office of Business Economics, U.S. Department of Commerce. (See Appendix A for a definition of regional subgroups.)

Within regions, PSUs were classified into five size-of-community (SOC) categories:

SOC 1 PSUs corresponding to the 13 largest SMSAs after adjusting the

population size to compensate for oversampling low-income metropolitan areas. These PSUs have selection probabilities so large that under NAEP allocation procedures they are certain to be included in the sample each year. These PSUs are designated as self-representing.

SOC 2 — PSUs corresponding to the remaining 57 SMSAs with over 500,000 population.

SOC 3 — PSUs corresponding to the remaining 162 SMSAs.

SOCs 4&5 — PSUs made up of non-SMSA counties. SOCs 4 and 5 are determined so that half of the remaining population (after adjustment for oversampling of rural areas) falls into each category. SOC 4 contains PSUs in which less than 60% of the residents are classified as rural.

The self-representing PSUs represented additional stratification, making an effective total of 17 (13 + 4) size-of-community strata. Each self-representing SMSA was divided further into geographical substrata or nonoverlapping replicates that constituted multiples of convenient work units for item administration. These multiple work units were included with the rest of the nonself-representing PSUs to form the pool from which first-stage sampling units were selected. To insure adequate representation, National Assessment doubled the sampling rate of low-income and rural areas.

In 1975-76, first-stage units were selected simultaneously for four consecutive assessment years (1975-76 through 1978-79), as were schools in the self-representing PSUs. The sample design required that every four years NAEP assess at least once in every state and not more than once in any school. There were 1,101 primary sampling units in the primary sampling frame for the four-year period, from which about 75 first-stage sampling units were selected each year.

Each year, within the primary strata public and private schools were listed and further stratified by the estimated number of youngsters in a school eligible at each age. Small schools were clustered until they formed a large enough group to respond to the same number of item booklets as the larger schools in a stratum. Schools or school clusters were selected without replacement, with probability proportional to the number of age-eligibles in the school or cluster of schools. Once schools were identified, districts were contacted to check for changes in grade range and for the existence of new schools. This information was used to revise probabilities of schools' selection.

In the third stage, students within each sample school were selected with equal probability and without replacement. The number of students selected was proportional to the number of age-eligibles, with oversampling in low-income and rural areas. In 1978-79, the size of groups selected for administrative sessions within schools was allowed to vary from 10 to 25 students. This enabled National Assessment to obtain desired sample sizes in schools having

characteristically low response rates. This feature also permitted last-minute modifications and adjustments to selection probabilities necessitated by enrollment changes.

The sampling procedures used in the 1969-70 and 1973-74 assessments differed somewhat from those used in the 1978-79 assessment (Final Report...Sampling and Weighting Activities..., 1980; Moore et al., 1974; Chromy and Horvitz, 1970). First, size measures for SMSAs, counties and urban areas in 1978-79 were based on 1970 census data, while those in 1969-70 were based on 1960 census data. Size measures in 1973-74 were based on 1960 census data and first-count data from the 1970 census.

Another difference occurred in the PSU sample design. In 1969-70, PSUs were stratified by region, size of community, a measure of socioeconomic status (SES) and geographic proximity. There was no requirement that all states be included in the sample. In 1973-74, the PSUs were stratified by region and size of community. In addition, the sample was constrained to include all states. In 1978-79, PSUs were stratified by region and size of community, with the constraint that each state must appear in the sample once very four years.

SOC categories were defined as shown previously in both the second and third assessments, but the number of SMSAs included in SOCs 1, 2 and 3 changed between 1973-74 and 1978-79. In 1973-74, there were 15 self-representing SMSA PSUs in SOC 1, 50 SMSAs in SOC 2 and 178 SMSAs in SOC 3.

The size-of-community stratifications used in 1973-74 and 1978-79 differed from those used in the 1969-70 assessment. In 1969-70, there were only four SOC stratifications. The first SOC category consisted of all central cities with overall population greater than 180,000; the second included the remainder of the SMSA containing the central city in SOC 1. SOC 3 consisted of the other SMSAs and all counties not included in SOC 1 or 2 that contained at least one city with a population over 15,000. SOC 4 consisted of all counties not included in SOC 1, 2 or 3.

In 1973-74 and 1978-79, oversampling of low-income metropolitan areas and extreme-rural areas was accomplished at the primary stage by increasing the estimated population size measures of PSUs containing these areas and then sampling with probabilities proportional to these adjusted size measures. In 1969-70, a poverty index was used to stratify PSUs into high- and low-SES stratifications. The sampling rates within these strata were then increased in order to achieve the desired oversampling. In the third stage, 12 students were selected for each administrative session.

Sampling of Essay Responses for Mechanics Scoring

The two essays at each age that had been given in all three writing assessments were designated for detailed mechanics and syntax scoring.

One of these essays at each age had been analyzed and released for publication following the 1973-74 assessment. For this essay, the scoring process involved replicating procedures used in 1973-74 to enable reporting of data from three assessments. Approximately 400 responses at an age were subsampled from the total number of papers collected in each assessment, allowing reporting of results for the nation and for males and females. Selection of sample responses from the first two assessments was done in 1973-74 using random sampling techniques. For those assessments, booklets containing the items were administered to groups of about 12 students. In sampling, booklets were ordered by student ID number and two booklets were chosen from each group administration using a table of random numbers. Following the third assessment, a systematic sample of one-fifth of the 1978-79 papers was drawn at each age. Papers were ordered by ID numbers within schools and every fifth response was chosen. For 9-year-olds, selection began with the fourth paper in the ordered list; for 13-year-olds, with the fifth paper; and for 17-year-olds, with the second paper.

The other essay at each age scored for mechanics and syntax was not analyzed and released until the third assessment. For these essays, an opportunity existed to revise the sample size, and it was decided to report data on blacks' and whites' performances as well as that of the nation and of males and females. Accordingly, a larger sample -- approximately 650 -- was drawn for each age with oversampling of blacks. A separate systematic sample was drawn for blacks and nonblacks. At ages 9 and 13, one of every two blacks and one of every five nonblacks was chosen. At age 17, due to a smaller number of responses per booklet, 5 of each 9 blacks and 3 of each 13 nonblacks were selected. Table 4 shows the sampling patterns used for these essays.

TABLE 4. Systematic Sampling Patterns for Mechanics and Syntax Scoring of Essays With Oversampling of Blacks

	Age 9	Age 13	Age 17
Blacks Select	1 of 2 2nd of each pair	1 of 2 1st of each pair	5 of 9 1, 3, 5, 7, 9 of each 9
Nonblacks Select	1 of 5 4th of each 5	1 of 5 2nd of each 5	3 of 13 3, 7, 11 of each 13

The numbers of students responding to each booklet containing writing exercises are shown in Chapter 5, and the sample sizes used in obtaining writing assessment results for the various scoring procedures for essay items of each assessment are detailed in Table 5.

TABLE 5. Sample Sizes Used for Reporting Writing Assessment Results

Age	Exercise	Analysis	No. in Sample 1969-70	1973-74	1978-79
9	"Kangaroo" (narration)	Holistic scoring	384	409	494
		Paragraph coherence	384	409	494
		Syntax and mechanics	384	409	494
13	"Describe" (descrip-tion)	Holistic scoring	395	420	536
		Paragraph coherence	395	420	536
		Syntax and mechanics	395	420	536
17	"Describe" (descrip-tion)	Holistic scoring	365	417	538
		Paragraph coherence	365	417	538
		Syntax and mechanics	365	417	538
9	"Fireflies" (narration)	Primary trait scoring	2,466	2,573	2,553
		Cohesion	2,466	2,573	2,553
		Syntax and mechanics	556	584	596
13	"Rainy Day" (expression)	Primary trait scoring	2,408	2,621	2,804
		Cohesion	2,408	2,621	2,804
		Syntax and mechanics	589	630	680
17	"Stork" (narration)	Primary trait scoring	2,073	2,281	2,748
		Cohesion	2,073	2,281	2,748
		Syntax and mechanics	594	596	722
9	"Goldfish" (expression)	Primary trait scoring		2,611	2,475
9	"Puppy Letter" (persuasion)	Primary trait scoring		2,643	2,494
9	"Poster Calendar" (explanation-business letter)	Primary trait scoring			2,492
13	"Loss" (expression)	Primary trait scoring		2,607	2,775
13	"Principal Letter" (persuasion)	Primary trait scoring		2,552	2,793

TABLE 5 (Continued). Sample Sizes Used for Reporting Writing Assessment Results

13	"Poster Calendar" (explanation-business letter)	Primary trait scoring		2,776
17	"Grape Peeler" (expression-humorous)	Primary trait scoring	2,283	2,765
17	"Recreation Center" (persuasion)	Primary trait scoring	2,308	2,784
17	"Electric Blanket" (explanation-business letter)	Primary trait scoring	2,776	2,781
13	Background questions			29,430
17	Background questions		34,211	26,651

The attitudinal questions, as well as some sentence-combining and cloze exercises were administered to all three age groups. All were administered for the first time in 1978-79, and all were administered and analyzed for full samples of 2,400-2,700 respondents at each age.

CHAPTER 5

DATA COLLECTION

National Assessment subcontracted data collection to the Research Triangle Institute and the Westinghouse DataScore Systems,[1] Iowa City, Iowa, for the 1969-70 writing assessment and to the Research Triangle Institute for the 1973-74 and 1978-79 assessments. A professional data collection staff was used rather than school personnel to minimize the burden on participating schools and to insure, insofar as possible, uniform administration conditions across the country (Final Report...In-School Field Operations..., 1979). In all three assessments, NAEP staff worked closely with the subcontractors to insure adherence to rigorous administrative standards.

Participation in the National Assessment is voluntary. NAEP makes every effort to encourage the schools selected in the sample to participate in the assessment, and National Assessment and Research Triangle Institute staffs have obtained high rates of school cooperation, as shown in Table 6 (Final Report...In-School Field Operations..., 1979, p. 39, Table 27). Student cooperation rates were also high. Percentages of the student sample covered are discussed in Appendix C. Table 7 shows the actual number of students who responded to the different exercise booklets at each age level in the 1978-79 assessment.

TABLE 6. School Cooperation Rates, 1978-79 Assessment

Age	Percent of Eligible Schools Participating in 1978-79 Assessment
9	90.4
13	90.9
17	92.9
Overall	91.3

[1]Formerly Measurement Research Center, Iowa City, Iowa.

TABLE 7. Number of Students Responding to Each Item Booklet in 1978-79 Assessment, by Age

Age 9		Age 13		Age 17	
Booklet	Number Responding	Booklet	Number Responding	Booklet	Number Responding
1	2,532	1	2,755	1	2,730
2	2,553	2	2,801	2	2,746
3	2,475	3	2,775	3	2,761
4	2,494	4	2,791	4	2,772
5	2,479	5	2,785	5	2,684
6	2,522	6	2,748	6	2,739
7	2,531	7	2,736	7	2,642
8	2,524	8	2,779	8	2,656
9	2,486	9	2,754	9	2,787
10	2,483+	10	2,758	10	2,697+
11	2,526+	11	2,751	11	2,628
		12	2,720+	12	2,628+
		13	2,757+	13	2,698+
				14	2,654+
Total	27,605		35,910		37,822

+There were no writing exercises included in booklets 10 and 11 at age 9, booklets 12 and 13 at age 13 and booklets 10, 12, 13 and 14 at age 17.

Each age group was assessed at approximately the same time of the school year in each of the three writing assessments: 13-year-olds were assessed in October-December, 9-year-olds in January-February and 17-year-olds in March-May. In 1978-79, booklets were administered to groups of 10-25 students, with each group responding to only one of the booklets for their age level. The groups varied in size depending on an estimate of the rate of nonresponse for a particular school. In 1969-70 and 1973-74, the planned session sizes were fixed at 12 students.

In each assessment, steps were taken to guarantee the anonymity of respondents. Students' names were listed with their booklet identification number so that scoring and processing personnel could go back to the school lists for data verification -- for instance, on background information -- if necessary. These lists did not leave the schools and were destroyed six months following the assessment in a school.

To provide information on respondents' backgrounds, school officials were asked to respond to a "principal's questionnaire," which included questions about the size and type of community served by the schools. In addition, in 1978-79, officials in 13- and 17-year-olds' schools were asked to respond to a

"supplementary principal's questionnaire," which asked about writing programs in the school. Students also provided information on their backgrounds through questions included in the item booklets. Samples of forms used to collect background information from students and school officials in the 1978-79 assessment appear in Appendix B. Changes from the two earlier assessments are described on the forms.

The assessment administrator coded each student's birth date, sex, grade, racial/ethnic classification and identification number on his or her booklet. Administrators made a visual racial/ethnic identification at the time each student turned in his or her booklet. During the 1978-79 assessment, six different racial classifications were used: white, black, Spanish heritage, American Indian or Alaskan native, Pacific Islander or Asian, and unclassified. If an administrator was unsure of a student's racial/ethnic group, the administrator referred to the student's name or listened to the student talk to make the identification. The assessment administrators did not ask students to give a racial identification for themselves; however, in 1979, 17-year-old students were asked to provide this information in one of the background questions included in the exercise booklet.

Sample sizes of the classifications American Indian or Alaskan native and Pacific Islander or Asian were too small to permit reporting for these groups. Also, results for the group classified as Spanish heritage cannot be reported for separate exercises, only for aggregates of exercises. Since the number of writing exercises was small, the primary analysis procedure for writing was to report results for individual exercises; thus, data for Hispanic students were not provided.

Following data collection, assessment administrators sent completed booklets to the scoring contractor, Westinghouse DataScore Systems, Iowa City, Iowa. Booklets were quality checked to verify that correct administrative procedures were followed by the field staff. Coded identification information was also checked for accuracy; inconsistencies that could not be reconciled were sent back to the assessment administrator to be checked against the list of student names and identification numbers retained by the school for six months following the assessment.

CHAPTER 6

SCORING

Trained scorers classified responses to essay and other open-ended items using detailed scoring guides, while responses to the few multiple-choice items in the writing assessment were read directly by optical-scanning machines. Scoring and computer recording of data were contracted to Westinghouse DataScore Systems, Iowa City, Iowa, for all three of the writing assessments.

National Assessment has found it most efficient to have scoring done by an outside contractor and to have the same contractor do both the machine scoring and the open-ended, or hand, scoring. Booklets to be scored do not have to be shipped to another location when different scoring methods are needed; in addition, the scoring contractor has a trained staff of scoring personnel that can be called upon and augmented when National Assessment conducts a major scoring effort.

Several types of scoring were used for the 1978-79 assessment. The procedures used with each exercise are listed in Chapter 4, Table 5. Each is briefly described below. Complete documentation of all exercises released after the third assessment of writing including scoring guides and sample responses is contained in Report 10-W-25, The Third Assessment of Writing: 1978-79 Released Exercise Set (1981). Readers desiring detailed information about the scoring procedures should consult the released exercise set, as well as Mullis (1980), Mullis and Mellon (1980) and Brown (1979).

Overview of Methods Used in Scoring Writing

Primary Trait System

Almost all of the essays and letters included in the writing assessment were evaluated using the primary trait system of scoring. This system describes a respondent's ability to choose and effectively carry out appropriate rhetorical strategies.

National Assessment uses a 4-point scale to describe levels of proficiency in the primary skill being assessed. Generally, level "1" indicates little or no evidence of the skill; level "2" indicates minimal evidence of the skill; level "3" indicates solid performance or competence; and level "4" is reserved for very good performance. Each essay task has a

scoring guide specifically tailored to the primary trait being measured. For example, explanatory writing should present facts in a clear and orderly fashion. At the lower end of the scoring categories would be a response with no information, clarity or logical order, while at the upper end would be an informative and unified presentation.

In addition to primary traits, NAEP gathered information about some secondary traits, such as tone or letter format. Such traits were generally categorized as present or not present -- for example, a letter either did or did not contain a greeting, closing, and so forth.

To insure accuracy of change measures, responses collected in different assessments were scored at the same time. Because scores given in an earlier assessment, even if the same scoring guides were used, might reflect slightly different judgment criteria, responses to unreleased writing tasks generally were not scored immediately following an assessment, but were held until the item was designated to be released. Then, responses from all assessments in which the item was included were scored together.

Holistic Scoring

Holistic scoring was used to rate one exercise at each age that had been included in all three writing assessments. When readers holistically score papers, they do not focus upon particular aspects of a paper such as mechanics, ideas or organization. Rather, they concentrate upon forming an overall impression of each paper relative to the other papers they have read. Since their primary task is to rank the papers from best to worst, this type of scoring tends to encourage a normal distribution. Thus, it is essential that sets of papers be scored together if results are to be compared because the distribution of holistic scores for any given subset of papers is highly dependent upon the entire set of papers with which it is scored. If two sets of papers are scored separately, the distribution of scores for each set may be quite different from what it would be if the two sets were scored together.

For each age, National Assessment randomly ordered papers collected in all three years into a single pool and conducted a single scoring session. This scoring was subcontracted to Edward White, University of California at San Bernardino. Readers had general guidelines describing four levels of quality in which papers could be categorized.

Since readers did not know which papers had been collected in which assessment, they necessarily applied the same criteria to all the papers. After the scoring session, the ratings were examined to determine whether groups of papers written in different years had received better or worse rankings than each other.

For holistic scoring, papers were read by only one scorer. Reliability of the scoring was checked by having a random 10% of the papers read by two scorers. The scorers agreed on 68% of the papers for age 9 and 79% of the papers for both ages 13 and 17.

Since the essays scored using the holistic method had been analyzed and released following the 1973-74 assessment, responses from the first two assessments were scored together at that time. Following the 1978-79 assessment, responses from the two previous assessments were rescored with the 1978-79 responses. It should be noted that scores previously reported for the first and second assessment papers will differ from the scores reported for those papers following the third assessment because they were reevaluated within a different total set of papers.

Coherence, Syntax and Mechanics

In addition to judging the overall quality of responses, National Assessment examined specific elements of coherence, mechanics and syntax; however, such scoring can be quite time-consuming and expensive. Thus, even though National Assessment recognizes the importance of specific and concrete information about these skills, it generally applies these types of scoring to only one or two writing tasks at an age level and leaves many of the analyses needed to supply detailed descriptive and diagnostic information to educational specialists and researchers.

The systems for evaluating coherence, syntax and mechanics have evolved considerably since the first writing assessment. In the 1969-70 assessment, NAEP scorers counted mechanical errors and characterized papers on the basis of correct and faulty constructions. For the 1973-74 assessment, National Assessment wished to describe level of complexity and sophistication of writing as well as tabulate mechanical errors. Thus, categorization of paragraph coherence and sentence types was added.

The paragraph coherence, mechanics and syntax scoring systems developed for the 1973-74 assessment were used with the subsamples of responses to the holistically scored exercises. Although these exercises were scored and released for publication following the second assessment, they were reassessed in 1978-79 to provide additional information about changes in performance across three points in time. A comparative subsample of responses from the third assessment was scored using the 1973-74 guidelines for paragraph coherence, mechanics and sentence types. (Details on the way in which the subsamples of responses were selected appear in Chapter 4 and sample sizes appear in Table 5.) Since these guidelines were more objective than those for the holistic scoring, it was decided that all first and second assessment responses did not have to be rescored with third assessment data. To insure that scoring was consistent, a reliability study was done, in which 10% of the responses from both the first and second assessments were rescored with third assessment responses. Agreement across the various categories averaged about 90%. However, for a few categories, standard errors used to calculate changes in performance over time were increased by an adjustment factor based on the amount of disagreement to insure that changes in performance would not be attributable to changes in scoring across time.

Following the 1973-74 assessment, NAEP responded to requests for more complete data on writing by conducting a T-unit analysis to provide more

accurate measures of syntactic development. Also, a new procedure, based on cohesive ties, was developed to describe coherence. These scoring methods, in addition to the primary trait system and mechanics scoring, were used to rate responses for one previously unreleased exercise at each age that had been administered in all three assessments. For each of these scoring procedures, papers from all three assessments were randomly ordered and scored together.

The term cohesion refers in general to the many ways words and ideas are linked together in writing to create a sense of wholeness and coherence. The cohesion scoring required readers to sort papers into groups representing four degrees of cohesiveness. Papers in the lowest group (level 1) display no or few connections between sentences and are loosely structured. Papers in the next group (level 2) display attempts to tie ideas together here or there but do not show any unifying structure. Cohesive papers (level 3) display gathering and ordering of details and ideas, and fully coherent papers (level 4) display a number of strategies and devices that bind the paper into a unified whole.

The T-unit analysis of syntax and the mechanics analysis were conducted only on a subsample of responses to these exercises. (See Chapter 4 for a description of the way in which the subsamples were selected and Table 5 for sample sizes.) Syntax refers to the ways in which words are put together to form phrases, clauses and sentences. The T-unit analysis involved breaking each paper up into its "T-units" (a T-unit is a main clause with all its attendant modifying words, phrases and dependent clauses) and examining the ways in which writers embedded information in T-units and joined T-units together. The mechanics analysis involved classifying and tabulating the kinds of errors writers made in sentence use, punctuation, spelling, and so forth.

Sentence combining exercises, assessed only in 1978-79, were hand scored using scoring guides that provided a description of the various combining strategies that could be used. The attitudinal questions, cloze paragraph exercises and most of the background questions were multiple-choice and were tabulated by machine.

Training of Scorers and Scorer Reliability

Westinghouse DataScore Systems and National Assessment staff worked together to train readers for all but the holistic evaluations discussed earlier. Different groups of scorers were used for the primary trait and cohesion scoring; however, training procedures were similar. In training sessions, readers were given the scoring guide for an item and responses that exemplified each scoring category. The reasons why responses were classified in particular categories were discussed; scorers' questions were answered and, if necessary, modifications were made to scoring guides. Readers then several papers and categorizations were discussed. This process until readers were familiar with the application of the scoring gui-

Using these training procedures and careful monitoring during scoring, National Assessment has achieved excellent scorer agreement. In scoring actual assessment data, most papers were scored by two readers for primary trait and cohesion. Any discrepancies in categorization were resolved by a third reader. The need for reconciliation occurred on less than 10% of the papers. Table 8 gives the interscorer agreement percentages for primary trait and cohesion scoring on exercises that required more than one scorer. Percentages are for

---

TABLE 8. Interscorer Percentages of Agreement for Primary Trait and Cohesion Scoring Conducted in 1978-79

	Percents of Scorer Agreement Primary Trait		
	1969-70 Papers	1973-74 Papers	1978-79 Papers
Age 9			
Fireflies (102012)	93.3	94.3	95.1
Goldfish (101006)		93.9	94.7
Puppy Letter (201001)		93.6	93.3
Age 13			
Rainy Day (102015)	93.1	94.1	92.4
Loss (101007)		91.2	91.8
Principal Letter (201006)		94.1	93.5
Age 17			
Stork (102016)	96.7	93.0	95.4
Grape Peeler (101015)		96.2	94.2
Rec Center (201007)		95.2	91.2

	Percents of Scorer Agreement Cohesion		
	1969-70 Papers	1973-74 Papers	1978-79 Papers
Age 9			
Fireflies (102012)	93.0	93.2	94.0
Age 13			
Rainy Day (102015)	91.5	93.5	90.9
Age 17			
Stork (102016)	93.1	94.0	94.2

the scoring conducted following the 1978-79 assessment of the randomly ordered pools of papers collected in the three different assessments. They are broken out by assessment year to show the uniformity of agreement.

Scoring procedures for two essay items ("Electric Blanket" and "Poster Calendar") were so straightforward that only one reader was used. Sentence combining exercises also required scoring by only one reader. In these cases, to check reliability a random 10% of the papers were read by two scorers. The percentages of agreement for the pairs of readers were about 99% for each task.

The mechanics and syntax analyses were conducted at the same time by a group of readers thoroughly trained in grammar, usage and linguistics. Training procedures were similar to those used for the other scoring methods, but lasted about four weeks due to the number and complexity of the categorizations. For mechanics and syntax categorizations, the problem was not so much with scorer disagreement on classifications as with ensuring that each paper had been fully read. To maintain reliable data when describing numerous and complex categories, papers were usually scored by at least three and sometimes by four or even five readers.

CHAPTER 7

DATA ANALYSIS

Measures of Achievement

The basic measure of achievement reported by National Assessment is the percentage of respondents answering a given item acceptably. This percentage is an estimate of the percentage of 9-, 13- or 17-year-olds who would respond acceptably to a given item if every 9-, 13- or 17-year-old in the country were assessed.

Percentages of acceptable responses are used because each item is designed as a separate measure of some aspect of an objective or subobjective. The purpose of National Assessment is to discover if more or fewer people are able to answer these items acceptably -- and thus meet the objectives -- over time.

Because of the nature of the writing assessment, exact definitions of "correct" or "acceptable" performance are sometimes debatable. Thus, for the primary trait system (PTS) and cohesion evaluations, percentages were reported for each response category. In addition, to help determine whether or not performance had changed over time, percentages of better papers from each assessment were totaled and compared. Primary trait percentages were computed for both levels 2, 3 and 4 (marginal or better) combined and levels 3 and 4 (good and very good) combined. Cohesion percentages were only computed for 3 and 4 combined, since level 2 papers were not considered really cohesive. Holistic scoring information was described in terms of percentages of papers at each score point and average holistic score for each assessment year.

Syntax and mechanics scoring data were analyzed to provide results about changes in the range of performance as well as the average performance. A frequency distribution was obtained for each characteristic (misspelled words per essay, awkward sentences per essay, subordination per T-unit, and so forth) for each set of papers at an age for each assessment year. Means, medians, quartiles and deciles were reported. The means provided information about changes over time in the average frequency with which characteristics occurred. Medians, quartiles and deciles described the changes in the distributions of each characteristic, giving the values below which 10%, 25%, 50%, 75% and 90% of the papers fell.

In addition to providing national results, National Assessment reported on the achievement of various subpopulations of interest. Groups are defined by region of the country, sex, race, size and type of community lived in,

level of parents' education and grade in school (see Appendix A for definitions of these groups). Results were also analyzed by frequency of writing done in school and instructional experiences with writing. (See Appendix B for questions used to obtain information about instructional experiences with writing.) Average syntax and mechanics characteristics were computed for good and poor papers (as defined by holistic, primary trait and cohesion ratings).

Procedures for estimating percentages responding in various ways and averages of different characteristics of the papers are dependent on the sample design. Each response by an individual was weighted and multiplied by an adjustment factor for nonresponse.[1] An estimate of the percentages of a particular age group that would have responded to a particular exercise in a particular way if the entire age group were assessed was defined as the weighted number of that type of response divided by the weighted number of all the responses. A similar ratio of weights was used to estimate percentages and averages for reporting groups or subpopulations of interest.[2]

The difference between percentages or averages for a reporting group and that of the entire age group (nation) on an exercise was used to describe the performance of any reporting group relative to the entire age group. This difference is a positive number if the group achieved a higher percentage or average than the entire age group and is a negative number if the group achieved a lower percentage or average. For example, a group performance of +1.8% indicates that the percentage of responses for the group is 1.8 percentage points higher than the national percentage of responses for that age level.

Increases or decreases in the percentages and averages betweeen two assessments were estimated by finding the difference between percentages or averages obtained from each assessment. A positive difference indicates an increase, and a negative difference indicates a decrease in those who responded in a particular way from one assessment to the next. These differences, or change measures, were used to indicate trends in achievement for an age level or subpopulation of interest. Changes in group differences from the nation between two assessments were used to indicate the relative trend of a group compared with the national trend of the age group.

In considering National Assessment's achievement measures, it is the differences in performance between assessment years, among groups and among ages that are most useful. By maintaining the same item or set of items in making these comparisons, we have a reasonable indicator of whether more or

---

[1] Appendix C discusses nonresponse in assessment samples.

[2] Following the 1978-79 assessment, a weighting-class adjustment procedure was used to smooth estimated population proportions across the 10 assessments conducted between 1969-70 and 1978-79. A discussion of this procedure is included in Appendix E.

fewer people know or can do something judged important.

Estimating Variability in Achievement Measures

National Assessment used a national probability sample at each age level to estimate the proportion of people who would complete an exercise in a particular way. The sample selected was one of a large number of all possible samples of the same size that could have been selected using the same sample design. Since an achievement measure computed from each of the possible samples would differ from one sample to another, the standard error of this statistic was used as a measure of the sampling variability among achievement measures from all possible samples. A standard error, based on one particular sample, served to estimate that sampling variability.

In the interest of sampling and cost efficiencies, National Assessment used a complex, stratified, multistage probability sample design. Typically, complex designs do not provide for unbiased or simple computation of sampling errors. A reasonably good approximation of standard error estimates of acceptable response percentages and averages was obtained by applying the jackknife procedure (Miller, 1964, pp. 1594-1705; Miller, 1968, pp. 567-82; Mosteller and Tukey, 1968) to first-stage sampling units within strata. Standard errors for achievement measures such as national percentages, group differences, means or mean differences for a particular assessment year were estimated directly, taking advantage of features of the jackknife procedure that are generic to all of these statistics.[3] Since samples for different assessments are independent, the standard errors of the differences in achievement measures between assessments can be estimated simply by the square root of the sum of squared standard errors from each of the assessments.

The standard error provides an estimate of sampling reliability for the achievement measures used by National Assessment. It is comprised of sampling error and other random error associated with the assessment of a specific item or set of items. Random error includes all possible nonsystematic error associated with administering specific exercises to specific students in specific situations. Random differences among scorers for open-ended items are also included in the standard errors.

National Assessment has adhered to a standard convention whereby differences between statistics are designated as statistically significant at the .05 level of significance. That is, differences in performance between assessment years or between a reporting group and the nation are highlighted with asterisks only if they are at least twice as large as their standard error. Differences this large would occur by chance in fewer than 5% of all possible replications of our sampling and data collection procedures for any particular reporting group or national estimates.

---

[3] See Appendix D for a more detailed description of National Assessment's computation of standard errors.

Controlling Nonrandom Errors

Systematic errors can be introduced at any stage of an assessment -- exercise development, preparation of exercise booklets, design or administrative procedures, field administration, scoring or analysis. These nonsampling, nonrandom errors rarely can be quantified, nor can the magnitude of the bias they introduce into the estimates be evaluated directly.

Systematic errors can be controlled in large part by employing uniform administrative and scoring procedures and requiring rigorous quality control in all phases of an assessment. If the systematic errors are the same from age to age or group to group, then the differences in percentages or mean percentages are measures with reduced bias because subtraction tends to cancel the effect of the systematic errors.

Similarly, the effect of systematic errors in different assessment years can be controlled by carefully replicating in the second assessment the procedures carried out in the first. Differences in achievement across assessment years will also be measures with reduced bias since subtraction will again tend to cancel systematic errors.

Although it is not possible for every condition or procedure to remain exactly the same between assessments conducted several years apart, National Assessment has made every effort to keep conditions as nearly the same as possible. Changes in procedures described in this report were judged to have a relatively minor impact.

CHAPTER 8

REPORTS ABOUT THE WRITING ASSESSMENTS

Each assessment generates a tremendous amount of information. To make the data as useful as possible to a variety of audiences, National Assessment provides several types of reports.

Reports

Since it is difficult and time-consuming to synthesize many discrete bits of data, National Assessment prepared reports about the writing assessments for the general public -- including parents, classroom teachers, school administrators and legislators -- that not only provided considerable data about each exercise but also synthesized and highlighted assessment results.

Although National Assessment does not interpret assessment results, it recognizes that data presented alone are often difficult to consider in perspective. Accordingly, National Assessment asked a group of writing educators to review and comment upon the results; their comments are included in the reports summarizing the writing assessment results.

Reports concerning changes in writing performance across the writing assessments are organized by age level. The following reports are available:

Writing Achievement, 1969-79: Results From the Third National Writing Assessment, Volume I — 17-Year-Olds, Report no. 10-W-01 (1980).

Writing Achievement, 1969-79: Results From the Third National Writing Assessment, Volume II — 13-Year-Olds, Report no. 10-W-02 (1980).

Writing Achievement, 1969-79: Results From the Third National Writing Assessment, Volume III — 9-Year-Olds, Report no. 10-W-03 (1980).

An additional report on the syntax used by 9-, 13- and 17-year-old students is planned for publication during 1981.

The following reports describing results from the second assessment are also available:

Writing Mechanics, 1969-1974: A Capsule Description of Changes in Writing Mechanics, Report no. 05-W-01 (1975).

Expressive Writing: Selected Results From the Second National Assessment of Writing, Report no. 05-W-02 (1976).

Explanatory and Persuasive Letter Writing, Report no. 05-W-03 (1977).

Write/Rewrite: An Assessment of Revision Skills, Report no. 05-W-04 (1977).

Objectives

A description of the 1973-74 and 1978-79 writing objectives and the procedures used in developing the objectives are available in Writing Objectives, Second Assessment (1972).

Exercises, Scoring Procedures and Data

For those wishing to use specific National Assessment items, National Assessment provides copies of released items, exercise documentation and scoring guides as well as information on scoring procedures and exercise-level data in microfiche. Materials that are or will be available include:

The Third Assessment of Writing, 1978-79, Released Exercise Set (1980) -- this loose-leaf set provides copies of all released writing exercises from the 1978-79 writing assessment, detailed scoring guides, exercise documentation (including timing and objective measured) and sample papers representing the various score points. Included with this set are the writing objectives and two papers, described below, which explain National Assessment's procedures for scoring essays.

Using the Primary Trait System for Evaluating Writing (Mullis, 1980) describes the rationale behind the primary trait system and how the system is applied.

Guidelines for Describing Other Aspects of Writing: Syntax, Cohesion and Mechanics (Mullis and Mellon, 1980) explains how syntax, cohesion and mechanics were scored in 1978-79 and presents summaries of the scoring guides used.

Data Appendix for the 1978-79 Writing Released Exercise Set (1981) -- this microfiched appendix will provide data on the exercises included in the Released Exercise Set. For each exercise, percentages of respondents answering in various response categories will be presented for the nation and for designated population groups.

For those desiring additional exercises, the exercises released following the second assessment are available in Writing Exercise Set (1976). Data for these items are found in Report no. 05-W-20, The Second National Assessment of Writing, New and Reassessed Exercises With Technical Information and Data (1978).

Public-Use Data Tape

For those who wish to perform their own analyses of National Assessment data, National Assessment will make available a data tape of respondent-level data for the 1978-79 writing assessment. To protect the confidentiality of the respondents, all identifying information (school, district, state) has been deleted. The tape includes documentation of exercises and is organized and documented in such a way that it can be used with standard statistical packages.

User Services

National Assessment provides some assistance to those wishing to use assessment items or to replicate assessment methodology. Those interested in receiving assistance should contact the National Assessment office.

APPENDIX A

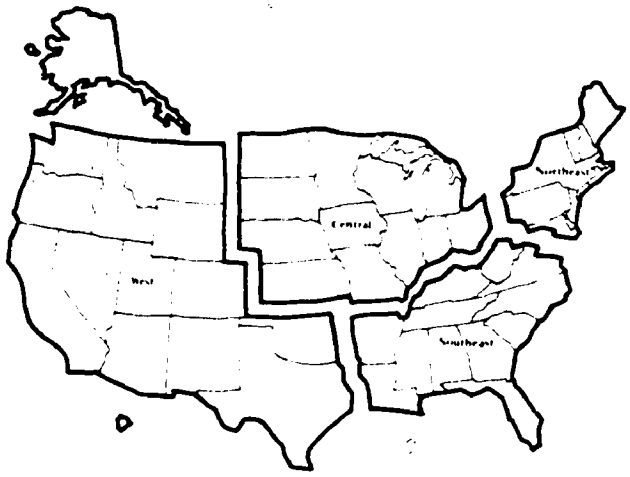
DEFINITIONS OF NATIONAL ASSESSMENT REPORTING GROUPS

In addition to reporting results for all 9-, 13- and 17-year-old students in the United States, National Assessment reports results for a number of population subgroups. Most of these subgroups were defined for all three writing assessments.

Reporting Variables

Region

The country has been divided into four regions: Northeast, Southeast, Central and West. States included in each region are shown on the following map.



Sex

Results are reported for males and females.

Race

Results are presented for blacks and whites.

Level of Parental Education

Three categories of parental-education levels are defined by National Assessment, based on students' reports. These categories are: (1) those whose parents did not graduate from high school, (2) those who have at least one parent who graduated from high school and (3) those who have at least one parent who has had some post high school education.

Type of Community

Three extreme community types of special interest are defined by an occupational profile of the area served by a school as well as by the size of the community in which the school is located.

Advantaged-urban (high-metro) communities. Students in this group attend schools in or around cities having a population greater than 200,000 where a high proportion of the residents are in professional or managerial positions.

Disadvantaged-urban (low-metro) communities. Students in this group attend schools in or around cities having a population greater than 200,000 where a relatively high proportion of the residents are on welfare or are not regularly employed.

Rural communities. Students in this group attend schools in areas with a population under 10,000 where many of the residents are farmers or farm workers.

This is the only reporting category that excludes a large number of respondents. About two-thirds do not fall into the classifications listed above. Results for the remaining two-thirds were not reported since their performance was similar to that of the nation.

Size of Community

Big cities. Students in this group attend schools within the city limits of cities having a 1970 census population over 200,000.

Fringes around big cities. Students in this group attend schools within metropolitan areas (1970 U.S. Bureau of the Census urbanized areas) served by cities having a population greater than 200,000, but outside the city limits.

Medium cities. Students in this group attend schools in cities having a population between 25,000 and 200,000, not classified in the fringes-around-big-cities category.

Small places. Students in this group attend schools in communities having a population less than 25,000, not classified in the fringes-around-big-cities category.

Grade in School

Results are categorized for 9-year-olds in the 3rd or 4th grade, 13-year-olds in the 7th or 8th grade, and 17-year-olds in the 10th, 11th or 12th grade.

Modal Grade by Region

Results are categorized for 9-, 13- and 17-year-old respondents in grades 4, 8 and 11, respectively, who live in the Northeastern, Southeastern, Central or Western regions of the country.

Modal Grade by Community Size

Results are categorized for 9-, 13- and 17-year-old respondents in grades 4, 8 and 11, respectively, who live in big cities, fringes around big cities, medium cities and small places.

Modal Grade by Sex

Results are categorized for 9-, 13- and 17-year-old males and females in grades 4, 8 and 11, respectively.

Instructional Experiences Related to Writing

Thirteen- and 17-year-old students were asked a number of questions regarding their instructional experiences with writing. Responses to these questions were used in analyzing writing achievement data. The questions employed to gather information about experiences follow and are found in Appendix B. These questions were given to 13-year-olds in the 1978-79 assessment and to 17-year-olds in both the 1973-74 and 1978-79 assessments.

Number of Papers Written: 13- and 17-Year-Olds

Students were asked how many reports they had written in the last 6 weeks as part of any school assignment. Five levels were defined: (1) no reports, (2) 1 report, (3) 2-4 reports, (4) 5-10 reports and (5) more than 10 reports (used at age 17 in 1973-74).

Instruction in Writing: 13- and 17-Year-Olds

Students were asked about the time spent in their English classes on instruction in writing. Three levels were defined: (1) little or none of the time, (2) about one-third of the time and (3) half or most of the time (used at age 17 in 1973-74).

Additional Remedial Writing Course: 17-Year-Olds

Seventeen-year-olds were asked if they had taken an additional remedial writing course. Two levels were defined: (1) yes and (2) no (used in 1973-74).

Additional Writing Course Other Than Remedial Writing: 17-Year-Olds

Seventeen-year-olds were asked if they had taken an additional creative writing course or any other additional course besides remedial writing. Two levels were defined: (1) yes and (2) no (used in 1973-74).

Encouraged Prewriting: 13- and 17-Year-Olds

Students were asked if their teachers encouraged them to make notes before writing and if their teachers encouraged them to make outlines before writing. Two levels were defined: (1) students who responded "usually" or "sometimes" to either or both questions and (2) students who responded "never" to both questions or who did not write papers.

Draft or Rewrite Papers Before Turning In: 13- and 17-Year-Olds

Students were asked if they wrote their papers more than once before turning them in to teachers. Three levels were defined: (1) students who responded "usually," (2) students who responded "sometimes" and (3) students who responded "never" or who did not write papers (used at age 17 in 1973-74).

Teacher Provides Written Suggestions on Papers: 13- and 17-Year-Olds

Students were asked if their returned papers had written suggestions from their teachers. Three levels were defined: (1) students who responded "usually," (2) students who responded "sometimes" and (3) students who responded "never" or had not written any papers (used at age 17 in 1973-74).

Teacher Discusses Papers: 13- and 17-Year-Olds

Students were asked if their teachers discussed their papers with them. Three levels were defined: (1) students who responded "usually," (2) students who responded "sometimes" and (3) students who responded "never" or had not written any papers.

Teacher Feedback: 13- and 17-Year-Olds

Using the information from students' responses about written suggestions from teachers and teacher discussions, two levels were defined: (1) students who responded "usually" or "sometimes" to either or both questions and (2) students who responded "never" to both questions or had not written any papers.

Work To Improve Papers After They Are Returned: 13- and 17-Year-Olds

Students were asked if they worked to improve papers after teachers

reviewed them. Three levels were defined: (1) students who responded "usually," (2) students who responded "sometimes" and (3) students who responded "never" or did not write any papers (used at age 17 in 1973-74).

Good and Poor Writers: 9-, 13- and 17-Year-Olds

For exercises scored for mechanics and syntax, three additional variables were analyzed: good and poor writers as defined by the (1) holistic, (2) primary trait or (3) cohesion evaluations. For each variable, the categories were defined as follows: (1) levels 1 and 2 combined (poor) and (2) levels 3 and 4 combined (good). These variables were only used in the analysis of syntax and mechanics characteristics. The variable(s) used was dependent on the type of scoring used in a particular exercise.

APPENDIX B

FORMS USED TO OBTAIN BACKGROUND INFORMATION

This appendix includes the forms used by National Assessment to collect background information from school officials and respondents for the 1978-79 assessment. Following are a listing and a brief decription of the forms included.

pp. 62–63 Writing Background Questionnaire for 17-Year-Olds -- provides information about frequency of writing assignments, additional writing courses, writing instruction and writing enjoyment.

SCHOOL PRINCIPAL'S QUESTIONNAIRE

This report is authorized by law (20 U.S.C. 1221 c-1). While you are not required to respond, your cooperation is needed to make the results of this survey comprehensive, accurate, and timely.

Primary Sampling Unit [ ][ ] School Number [ ][ ][ ][ ][ ][ ][ ][ ][ ]

Age Group(s) 9 13 17

PLEASE PRINT

Name of School _____

Address of School _____
(Street)

_____ _____ _____
(City) (State) (Zip Code)

Name of School Principal _____

Name and title of person completing the form if other than school principal

Name _____ Title _____

1. What is your best estimate of the current enrollment and the average daily attendance by grade of your school (1978-79 school year)? (Enter zeros for grades not served by your school.)

Grade	K	1	2	3	4	5	6	7	8	9	10	11	12
Enrollment													
Average Daily Attendance													

2. Approximately what percentage of the students attending your school live in each of the following areas?

_____% A In a rural area (less than 2,500)

_____% B In a town of 2,500 to 10,000

_____% C In a town of 10,000 or more

_____
100%

(Items A-C should add to 100%)

3. Approximately what percentage of the students attending your school are children of

_____% A Professional or managerial personnel

_____% B Sales, clerical, technical or skilled workers

_____% C Factory or other blue collar workers

_____% D Farm workers

_____% E Persons not regularly employed

_____% F Persons on welfare

_____ (Items A-F should add to 100%)
100%

4. Approximately what percentage of the students attending your school are

_____% A American Indian or Alaskan Native

_____% B Asian or Pacific Islander

_____% C Hispanic, regardless of race (Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin)

_____% D Black and not Hispanic

_____% E White and not Hispanic

_____ (Items A-E should add to 100%)
100%

5. Does your school qualify for ESEA Title I assistance?

_____ Yes - If Yes, approximately what number of students qualify for and what number of students are receiving ESEA Title I assistance?

_____ Approximate number of students qualifying for ESEA Title I assistance

_____ Approximate number of students receiving ESEA Title I assistance

_____ No

THANK YOU FOR YOUR COOPERATION

Supplementary Principal's Questionnaire

Instructions: The purpose of this questionnaire is to provide additional information which will be used in the analyses of NAEP data. Darken the appropriate ovals with a soft lead pencil. If you have questions about any of the following items, please contact the National Assessment district supervisor. Thank you for your cooperation.

1. Does your school:

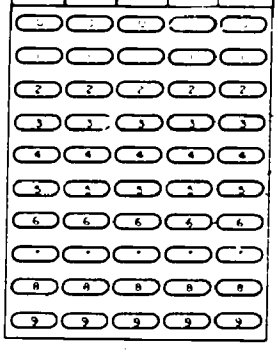
		Yes	No
A.	Provide in-service training, instructional workshops, or professional improvement programs on how to teach writing?	⬭	⬭
B.	Have any form of individualized writing instruction for students, such as a writing lab, a remedial writing program, or writing tutors?	⬭	⬭
C.	Have an active departmental committee on writing?	⬭	⬭
D.	Assign fewer classes to teachers of writing?	⬭	⬭
E.	Provide assistants to writing teachers to help read student essays and papers (such as adults in your community, teachers in other areas in your school, teaching aids, or outstanding student writers)?	⬭	⬭

2. What is the average size of the language arts (general English, literature or grammar) classes in your school?

- ⬭ 0 to 12 students
- ⬭ 13 to 18 students
- ⬭ 19 to 26 students
- ⬭ 27 to 31 students
- ⬭ 32 to 36 students
- ⬭ Over 36 students

THANK YOU FOR YOUR COOPERATION

P.S.U. and School Number

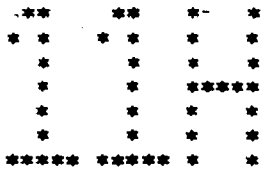


School Name: _____

Example Package Cover Sheet

O.M.B. No. 051-R1204
Approval Expires 9-30-81

PACKAGE NO.



NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
A PROJECT OF THE EDUCATION COMMISSION OF THE STATES

This report is authorized by law (20 U.S.C. 1221 e-1). While you are not required to respond, your cooperation is needed to make the results of this survey comprehensive, accurate and timely.

PACKAGE NUMBER	
1	1

I. D. NUMBER			

3 4 5 6 7 8

N
P
NR
NE

* Grade 5 or lower

** Grade 9 or higher

YEAR 10
AGE CLASS 2
GROUP

NCES Form Nos 2371 — 10, 57, 59

Directions to Exercise Administrators for Coding Package Cover Sheet

Codes To Be Used in Columns 3-8

Package Coding Column	Administration Schedule Column	Item	Code
3	3	Grade	Two digits: 8th grade = 08 Ungraded class = 98 Special education class = 99
4	4	Sex	1 = Male 2 = Female
5	5	Birthdate	Month and last two digits of year: May 1965 = 0565
6	6	Race	1 = W(White) 2 = B(Black) 3 = S(Spanish Heritage) 4 = I(American Indian or Alaskan Native) 5 A(Asian or Pacific Islander) 6 = U(Unclassified)
7	--	EA Number	Two digits; number recorded on the front cover of your manual
8	--	PSU and School Numbers	Five digits; First two = PSU Number; Last three = School Number; as shown on the Administration Schedule

Standard Background Information Form for 9-Year-Olds

1. Does your family get a newspaper regularly?

   ⊂⊃ Yes ⊂⊃ No ⊂⊃ I don't know.

2. Does your family get any magazines regularly?

   ⊂⊃ Yes ⊂⊃ No ⊂⊃ I don't know.

3. Are there more than 25 books in your home?

   ⊂⊃ Yes ⊂⊃ No ⊂⊃ I don't know.

4. Is there an encyclopedia in your home?

   ⊂⊃ Yes ⊂⊃ No ⊂⊃ I don't know.

5. How much school did your father complete?
   (FILL IN THE ONE OVAL which best shows how much school your father completed.)

   ⊂⊃ Did not complete the 8th grade

   ⊂⊃ Completed the 8th grade, but did not go to high school

   ⊂⊃ Went to high school, but did not graduate from high school

   ⊂⊃ Graduated from high school

   ⊂⊃ Some education after graduation from high school

   ⊂⊃ I don't know.

6. Did your father graduate from a college or university?

   ⊂⊃ Yes ⊂⊃ No ⊂⊃ I don't know.

7. How much school did your mother complete?
(FILL IN THE ONE OVAL which best shows how much school your mother completed.)

- ⬭ Did not complete the 8th grade
- ⬭ Completed the 8th grade, but did not go to high school
- ⬭ Went to high school, but did not graduate from high school
- ⬭ Graduated from high school
- ⬭ Some education after graduation from high school
- ⬭ I don't know.

8. Did your mother graduate from a college or university?

⬭ Yes ⬭ No ⬭ I don't know.

1. Does your family get a newspaper regularly?

   ⬭ Yes ⬭ No ⬭ I don't know.

2. Does your family get any magazines regularly?

   ⬭ Yes ⬭ No ⬭ I don't know.

3. Are there more than 25 books in your home?

   ⬭ Yes ⬭ No ⬭ I don't know.

4. Is there an encyclopedia in your home?

   ⬭ Yes ⬭ No ⬭ I don't know.

5. How much school did your father complete?
   (FILL IN THE ONE OVAL which best shows how much school your father completed.)

   ⬭ Did not complete the 8th grade
   ⬭ Completed the 8th grade, but did not go to high school
   ⬭ Went to high school, but did not graduate from high school
   ⬭ Graduated from high school
   ⬭ Some education after graduation from high school

   ⬭ I don't know.

6. Did your father graduate from a college or university?

   ⬭ Yes ⬭ No ⬭ I don't know.

7. How much school did your mother complete?
   (FILL IN THE ONE OVAL which best shows how much school your mother completed.)

   ⬭ Did not complete the 8th grade
   ⬭ Completed the 8th grade, but did not go to high school
   ⬭ Went to high school, but did not graduate from high school
   ⬭ Graduated from high school
   ⬭ Some education after graduation from high school

   ⬭ I don't know.

8. Did your mother graduate from a college or university?

   ⬭ Yes ⬭ No ⬭ I don't know.

9. Where did you live on your ninth birthday?

⊂⊃ In the United States (Please specify the state or territory.)

⊂⊃ Outside the United States (Please specify the country.)

⊂⊃ I don't know.

A. How much school did your father complete?
(FILL IN THE ONE OVAL which best shows how much school your father completed.)

- ⊂⊃ Did not complete the 8th grade
- ⊂⊃ Completed the 8th grade, but did not go to high school
- ⊂⊃ Went to high school, but did not graduate from high school
- ⊂⊃ Graduated from high school
- ⊂⊃ Some education after graduation from high school

⊂⊃ I don't know.

B. Did your father graduate from a college or university?

⊂⊃ Yes ⊂⊃ No ⊂⊃ I don't know.

C. How much school did your mother complete?
(FILL IN THE ONE OVAL which best shows how much school your mother completed.)

- ⊂⊃ Did not complete the 8th grade
- ⊂⊃ Completed the 8th grade, but did not go to high school
- ⊂⊃ Went to high school, but did not graduate from high school
- ⊂⊃ Graduated from high school
- ⊂⊃ Some education after graduation from high school

⊂⊃ I don't know.

D. Did your mother graduate from a college or university?

⊂⊃ Yes ⊂⊃ No ⊂⊃ I don't know.

E. Where did you live on your ninth birthday?

⊂⊃ In the United States (Please specify the state or territory.)

_____

⊂⊃ Outside of the United States (Please specify the country.)

_____

⊂⊃ I don't know.

F. Where did you live on your thirteenth birthday?

⊂⊃ In the United States (Please specify the state or territory.)

_____

⊂⊃ Outside of the United States (Please specify the country.)

_____

⊂⊃ I don't know.

1. Which of the following does your family have at home? (Fill in one oval on each line.)

		Have	Do not have
A.	Newspaper received regularly	⊂⊃	⊂⊃
B.	Magazines received regularly	⊂⊃	⊂⊃
C.	More than 25 books	⊂⊃	⊂⊃
D.	Encyclopedia	⊂⊃	⊂⊃
E.	Dictionary	⊂⊃	⊂⊃
F.	Record player	⊂⊃	⊂⊃
G.	Tape recorder or cassette player	⊂⊃	⊂⊃
H.	Typewriter	⊂⊃	⊂⊃
I.	Vacuum cleaner	⊂⊃	⊂⊃
J.	Electric dishwasher	⊂⊃	⊂⊃
K.	Two or more cars or trucks that run	⊂⊃	⊂⊃

2. How much time did you spend on homework yesterday?

⊂⊃ No homework was assigned
⊂⊃ I had homework but didn't do it
⊂⊃ Less than one hour
⊂⊃ Between 1 and 2 hours
⊂⊃ More than 2 hours

3. How many different schools have you attended since you started the first grade?

- ⬭ 1 to 3 schools
- ⬭ 4 to 6 schools
- ⬭ 7 to 9 schools
- ⬭ 10 or more schools

4. How long have you lived in the community in which you now live?

- ⬭ All my life
- ⬭ 10 or more years but not all my life
- ⬭ 5 to 9 years
- ⬭ 2 to 4 years
- ⬭ 1 year
- ⬭ Less than 1 year

5. How much television did you watch yesterday?

⬭ None	⬭ 2 hours	⬭ 5 hours
⬭ 1 hour or less	⬭ 3 hours	⬭ 6 hours or more
⬭ 1 hour	⬭ 4 hours	

6. Is English the language spoken most often in your home?

⬭ Yes ⬭ No

7. Is a language other than English spoken in your home?

⬭ Often ⬭ Sometimes ⬭ Never

8. How many brothers or sisters do you have who are <u>older</u> than you?

None	1	2	3	4	5	6 or more
⬭	⬭	⬭	⬭	⬭	⬭	⬭

9. How many brothers or sisters do you have who are <u>younger</u> than you?

None	1	2	3	4	5	6 or more
⬭	⬭	⬭	⬭	⬭	⬭	⬭

10. A. What is your racial background?

- ⬭ American Indian or Alaskan Native
- ⬭ Asian or Pacific Islander
- ⬭ Black
- ⬭ White
- ⬭ Other (Please specify) _____

B. Is your ethnic heritage Hispanic (such as Mexican, Puerto Rican, Cuban, Central or South American or other Spanish Culture or origin)?

⬭ Yes ⬭ No

11. How often has each of the following been used in the courses you are taking this year? (Fill in one oval on each line.)

	Never	Seldom	Fairly Often	Frequently
A. Listening to the teacher's lecture	⬭	⬭	⬭	⬭
B. Participating in student-centered discussions	⬭	⬭	⬭	⬭
C. Working on a project or in a laboratory	⬭	⬭	⬭	⬭
D. Writing essays, themes, poetry, stories	⬭	⬭	⬭	⬭
E. Going on field trips	⬭	⬭	⬭	⬭
F. Having individualized instruction (small groups or one-to-one with a teacher)	⬭	⬭	⬭	⬭
G. Using teaching machines or computer-assisted instruction	⬭	⬭	⬭	⬭
H. Watching television lectures	⬭	⬭	⬭	⬭
I. Studying from textbooks	⬭	⬭	⬭	⬭
J. Library or media center assignments	⬭	⬭	⬭	⬭

A. How many reports and essays have you written during the last six weeks as part of any school assignment?

B. In the general English, literature or grammar classes you have taken during the past two years, about what part of the class time was spent on instruction in how to write reports and essays?

- ⊃ None of the time
- ⊃ Little of the time
- ⊃ About one-third of the time
- ⊃ About one-half of the time
- ⊃ Most of the time

C. Are you encouraged to jot down ideas and make notes about the topic of your paper before you write it?

Usually	Sometimes	Never	I haven't written any papers.
⊃	⊃	⊃	⊃

D. Are you encouraged to make outlines of your papers before you write them?

Usually	Sometimes	Never	I haven't written any papers.
⊃	⊃	⊃	⊃

PLEASE CONTINUE ON THE NEXT PAGE

E. Do you write a paper more than once before you turn it in to your teachers?

Usually ⬭ Sometimes ⬭ Never ⬭ I haven't written any papers. ⬭

F. When your papers are returned, do they have written suggestions on how to improve your writing?

Usually ⬭ Sometimes ⬭ Never ⬭ I haven't written any papers. ⬭

G. When your papers are returned, do your teachers discuss them with you?

Usually ⬭ Sometimes ⬭ Never ⬭ I haven't written any papers. ⬭

H. After your papers are returned, do you work on the paper again to improve it?

Usually ⬭ Sometimes ⬭ Never ⬭ I haven't written any papers. ⬭

I. Do you enjoy working on writing assignments?

Usually ⬭ Sometimes ⬭ Never ⬭ I haven't written any papers. ⬭

DO NOT CONTINUE UNTIL TOLD TO DO SO.

A. How many reports and essays have you written during the last six weeks as part of any school assignment? _____

B. In the general English, literature or grammar classes you have taken during the past two years, about what part of the class time was spent on instruction in how to write reports and essays?

- ⬭ None of the time
- ⬭ Little of the time
- ⬭ About one-third of the time
- ⬭ About one-half of the time
- ⬭ Most of the time

C. In addition to the general English, literature or grammar classes you have taken during the last two years, have you had or are you now taking any of the following courses concerned with how to write?

	Yes	No	I don't know.
1. Creative writing course	⬭	⬭	⬭
2. Remedial writing course	⬭	⬭	⬭
3. Other writing course (If other, please specify.) _____	⬭	⬭	⬭

D. Are you encouraged to jot down ideas and make notes about the topic of your paper before you write it?

Usually ⬭ Sometimes ⬭ Never ⬭ I haven't written any papers. ⬭

E. Are you encouraged to make outlines of your papers before you write them?

Usually ⬭ Sometimes ⬭ Never ⬭ I haven't written any papers. ⬭

PLEASE CONTINUE ON THE NEXT PAGE

F. Do you write a paper more than once before you turn it in to your teachers?

Usually ⊂⊃ Sometimes ⊂⊃ Never ⊂⊃ I haven't written any papers. ⊂⊃

G. When your papers are returned, do they have written suggestions on how to improve your writing?

Usually ⊂⊃ Sometimes ⊂⊃ Never ⊂⊃ I haven't written any papers. ⊂⊃

H. When your papers are returned, do your teachers discuss them with you?

Usually ⊂⊃ Sometimes ⊂⊃ Never ⊂⊃ I haven't written any papers. ⊂⊃

I. After your papers are returned, do you work on the paper again to improve it?

Usually ⊂⊃ Sometimes ⊂⊃ Never ⊂⊃ I haven't written any papers. ⊂⊃

J. Do you enjoy working on writing assignments?

Usually ⊂⊃ Sometimes ⊂⊃ Never ⊂⊃ I haven't written any papers. ⊂⊃

Questions 4, 5, 8, 9 and 10 were not administered in 1974; however, the remaining questions were asked in both the 1973-74 and 1978-79 assessments.

STOP

DO NOT CONTINUE UNTIL TOLD TO DO SO.

APPENDIX C

RESPONSE RATES FOR ASSESSMENT SAMPLES

Table C-1 shows the response rates for students assessed in 1969-70, 1973-74 and 1978-79. In the 1969-70 and 1973-74 assessments the same general procedures were used to insure sample coverage. For each of the three age groups, 12 students and 4 alternates were selected for each assessment session. If all 12 students appeared for the session, then the alternates were dismissed. Otherwise, enough alternates were selected to bring the size of the group to as many students as possible up to a total of 12. If the group assessed numbered between 8 and 12 students, then the administration was considered complete. If the final total was not at least a quorum of 8, a second and sometimes a third make-up session was held. The percentages in Table C-1 are based on the numbers of students assessed from the original groups of 12 selected and do not reflect the use of alternates.

For the 1978-79 assessment, slightly different procedures were used. The number of students selected for each administrative session varied from 16-25 students depending on previous response rates obtained from schools in similar communities. No alternates were selected. The quorum size needed to consider an administrative session complete varied according to the number of students selected. Since nonresponse rates have always been relatively small for ages 9 and 13, the make-up or follow-up procedures used in 1978-79 for those ages were similar to the ones used for the first two writing assessments. If a quorum was not obtained at the first administrative session, a second and sometimes a third make-up session was held. At age 17, in the 1978-79 assessment, follow-up procedures were conducted on a school, rather than a session, basis. If a school had an overall response rate of less than 75%, then all nonrespondents in the school were contacted for one or two follow-up sessions. These follow-up procedures for 17-year-olds provided sample coverage similar to that obtained at ages 9 and 13.

Since response rates at age 17 have always been somewhat lower than at the other two ages, the Research Triangle Institute (RTI), Raleigh, North Carolina, was asked to conduct a special study of nonrespondents during the 1972-73 assessment of science and mathematics. The results (Kalsbeek et al., 1975; Rogers et al., 1977) indicated that about 80% of the total nonrespondent group did not appear at the assessment sessions because of conflicting school activities or illness. The remaining 20% did not seem to be available. They attended school infrequently, if at all (for practical purposes, they had dropped out), or they had moved out of the school attendance area. In either case, these students probably should not have been listed in the in-school population of eligibles.

TABLE C-1. Number of Students Assessed and Response Rates, by Age and Assessment Year

Year	Age	Type of Administration+	Number of Packages	Total Number of Students Assessed	Average Number Assessed Per Package	Response Rates in Percent
1969-70	9	G	8	19,468	2,434	88.0
		I	2	3,713	1,856	89.1
	13	G	9	21,696	2,411	85.6
		I	3	5,568	1,856	87.2
	17	G	11	22,913	2,083	74.5
		I	2	3,328	1,664	71.2
1973-74	9	G	10	26,503	2,650	90.4
		I	1	2,233	2,233	89.1
	13	G	14	36,080	2,577	86.5
		I	1	2,239	2,239	88.6
	17	G	16	36,709	2,294	75.2
		I	1	2,163	2,163	79.3
1978-79	9	G	11	27,605	2,510	87.8
	13	G	13	35,910	2,762	84.9
	17	G	14	37,822	2,702	81.4

+In the 1969-70 and 1973-74 assessments, some booklets were administered to individual students using an interview mode (I). The other booklets were administered to groups of students (G).

Tables published in previous National Assessment reports showing response rates for age 17 generally contain percentages adjusted to account for those 17-year-olds listed, but not attending school. But, since National Assessment has not had the resources to replicate the RTI study in recent assessments, the 20% figure used as a basis for adjusting these percentages may be outdated and thus, the percentages given in Table C-1 have not been adjusted. It seems likely that despite efforts to update the lists of eligibles, these lists still contain some percentage of students who have in effect left the schools. Thus, the percentages listed for age 17 are probably underestimates of the actual response rates for 17-year-olds attending school.

APPENDIX D

COMPUTATION OF MEASURES OF ACHIEVEMENT, CHANGES IN ACHIEVEMENT AND STANDARD ERRORS

Several measures of achievement that National Assessment uses in its reports are described in Chapter 7 of this document. The sample design, as described in Chapter 4, is a complex, deeply stratified, multistage probability sample design. Measures of achievement are obtained by weighting individual responses appropriately. Reasonably good approximation of standard error estimates of these achievement measures can be obtained by applying the jackknife procedure to first-stage sampling units (FSUs) within strata, using the method of successive differences and accumulating across strata.

In this section, the measures of achievement are first defined in algebraic form, followed by a description of the jackknife method that National Assessment uses to estimate the standard errors of achievement measures.

Measures of Achievement

Based on the sample design, a weight is assigned to every individual who responds to an exercise administered in an assessment. The weight is the reciprocal of the probability of selecting a particular individual to take a particular exercise with adjustment for nonresponse. Since the probabilities of selection are based on an estimated number of people in the target age population, the weight for an individual estimates the number of similar people that individual represents in the age population. As explained in Appendix E, the weights are adjusted to reflect information from previous assessments on population distributions.

A sum of the weights for all individuals at an age level responding to an exercise is an estimate of the total number of people in that age population. A sum of weights for all individuals at an age responding correctly to an exercise is an estimate of the number of people in the age population who would be able to respond correctly if the entire population were assessed. These concepts also apply to any reporting group (e.g., defined by region, sex, and so on) and category of response (e.g., correct, incorrect and "I don't know").

Let W^{e}_{ihk} = sum of weights for respondents to exercise e who are in reporting subgroup i and who are in the kth replicate of the hth sampling stratum, and

C^{ej}_{ihk} = sum of weights for respondents to exercise e who are in reporting subgroup i, who are in the kth FSU of the hth sampling stratum and who selected response category j (e.g., correct foil) for the exercise.

Note that $W^{e}_{ihk} = \sum_{j} C^{ej}_{ihk}$.

Then summing k over the n_h sample FSUs in the stratum h, and summing over the H sampling strata, $W^{e}_{i++} = \sum_{h=1}^{H} \sum^{n_h}_{k=1} W^{e}_{ihk}$ estimates the number of eligibles in the population who are in subgroup i.

Similarly, $C^{ej}_{i++} = \sum_{h=1}^{H} \sum^{n_h}_{k=1} C^{ej}_{ihk}$

estimates the number of eligibles in the population who are in subgroup i and who would select response category j for exercise e.

An estimate of the proportion of the eligibles in the age population in group i who would select response category j on exercise e is:

(1) $P^{ej}_{i} = C^{ej}_{i++} / W^{e}_{i++}$.

In the special case where the proportion of all age eligibles who would select response category j on exercise e is estimated, the index A (for ALL) will be used in place of i as follows:

(2) $P^{ej}_{A} = C^{ej}_{A++} / W^{e}_{A++}$.

In National Assessment reports, the proportion in (1) multiplied by 100 is called the group percentage, and the proportion in (2) multiplied by 100 is called the national percentage. The difference between the proportion in subgroup i who would select category j on exercise e and the proportion in the nation is denoted by:

(3) $\Delta P^{ej}_{i} = P^{ej}_{i} - P^{ej}_{A}$.

National Assessment also reports the arithmetic mean of the percentage of correct responses over sets of exercises corresponding to the measures in (1), (2) and (3). These means are taken over the set of all exercises or a subset of exercises classified by a reporting topic or content objective. The mean percentages of correct responses taken over m exercises in some set of

exercises corresponding to measures (1), (2) and (3) are, respectively:

$$(4)\quad \bar{P}_i = \frac{1}{m}\sum_e C^e_{i++} / W^e_{i++},$$

$$(5)\quad \bar{P}_A = \frac{1}{m}\sum_e C^e_{A++} / W^e_{i++} \text{ and}$$

$$(6)\quad \overline{1-P}_i + \bar{P}_i - \bar{P}_A.$$

Note that response category subscript j has been suppressed since the means are understood to be taken over the correct response category for each exercise.

Each of these six achievement measures is computed and routinely used in reports describing achievement data for any assessment. The simple difference in these measures between two assessments of the same exercise (or sets of exercises) provides six measures of change in achievement that are routinely used in National Assessment's change reports. The next section describes how standard errors are estimated for the 12 statistics used in NAEP reports.

Computation of Standard Errors

In order to obtain an approximate measure of the sampling variability in the statistics (1) through (6), a jackknife replication procedure for estimating the sampling variance of nonlinear statistics from complex, multistage samples was tailored to National Assessment's sample design. Miller (1968, 1974) and Mosteller and Tukey (1977) provide information about the jackknife technique, while Folsom (1977) describes how the procedure is used in estimating standard errors for National Assessment's sample design.

To demonstrate the computational aspects of this technique, consider estimating the variance of the statistic in (1) -- the proportion of age-eligibles in subgroup i who would select response category j on exercise e.

This statistic is based on the data from all the n_h FSUs in the H strata. Let P^{ej}_{i-hk} be defined as a replication estimate of P^{ej}_i and constructed from all the replicates excluding the data from FSU k in stratum h. These replication estimates are computed as if the excluded FSU had not responded,

and a reasonable nonresponse adjustment is used to replace the data in FSU hk in estimating p_i^{ej} . Several choices for replacing the data in FSU hk are available. In order to obtain a convenient and computationally efficient algorithm for approximating standard errors, National Assessment replaces c_{ihk}^{ej} and w_{ihk}^{e} from the hkth FSU with corresponding sums from another paired FSU in the same stratum. The replicate estimate is then computed. The replicate estimates to be used in the calculations are determined by arranging all the FSUs in each stratum into successive pairs. That is, FSU 1 is paired with FSU 2, FSU 2 with FSU 3, 3 with 4, ... (n_h-1) with n_h and FSU n_h with FSU 1.

The contribution to the variance of p_i^{ej} by each pair of FSUs is the change in the value of the statistic incurred by replacing the data from each FSU in the pair with the data from the other FSU in the pair and recomputing p_i^{ej} in the usual way. This produces two replicate estimates. Squaring the difference between these replicate estimates and then dividing by eight measures the contribution of this pair of FSUs to the total variance. The sum of these contributions over all n_h successive pairs in the stratum is the contribution by stratum h to the total variance. The square root of the sum of the H stratum contributions is the estimate of the standard error of p_i^{ej} .

Algebraically, the two replicate estimates for the pair k, k+1 (where $k=1,\ldots n_h$ and $n_h+1=1$) are:

$$(7)\quad p_{i-hk}^{ej} = \frac{c_{i++}^{ej} - c_{ihk}^{ej} + c_{ih(k+1)}^{ej}}{w_{i++}^{ej} - w_{ihk}^{ej} + w_{ih(k+1)}^{ej}}$$

and

$$(8)\quad p_{i-h(k+1)}^{ej} = \frac{c_{i++}^{ej} - c_{ih(k+1)}^{ej} + c_{ihk}^{ej}}{w_{i++}^{ej} - w_{ih(k+1)}^{ej} + w_{ihk}^{ej}}.$$

The contribution to the total variance from stratum h is:

$$(9)\quad \text{var}\ (P_{ih}^{ej}) = \frac{1}{d}\sum_{k}^{n_h}\left(P_{i-hk}^{ej} - P_{i-n(k+1)}^{ej}\right)^2.$$

And finally, an estimate of the standard error of P_i^{ej} is:

$$(10)\quad SE\ (P_i^{ej}) = \left(\sum_{h}^{H} \text{var}\ P_{ih}^{ej}\right)^{\frac{1}{2}}.$$

Multiplying P_i^{ej} by 100 yields the percentage of response to category j. Multiplying $SE(P_i^{ej})$ by 100 yields the corresponding estimated standard error of the percentage.

In general, the jackknifed standard errors of the proportion estimates will be larger than the simple random sampling formula $(PQ/N)^{\frac{1}{2}}$, where $P=P_i^{ej}$, $Q=1-P$ and N is the number of sampled respondents in subgroup i who took the exercise. The larger size of $SE\ (P_i^{ej})$ reflects mainly the loss of precision due to cluster-sampling of schools and students. The standard errors for the achievement measures (2) through (6) are computed through a series of steps analogous to those followed in computing $SE\ (P_i^{ej})$.

The standard errors for the differences between two assessments for any of the achievement measures (1) through (6) are computed as the square root of the sum of the squared standard errors from each of the separate assessments.

The size of the standard errors depends largely not only on the number of FSUs and schools included in the sample, but also on the number of respondents in each of the reporting groups. Table D-1 shows the average number of students responding to an exercise booklet for each of the reporting groups for each age for each assessment year. Table D-2 shows National Assessment's current estimates of the proportions of students in each reporting group at each age.

TABLE D-1. Average Number of Respondents in Reporting Groups Taking an Item Booklet, by Age and Assessment Year+

	Age 9			Age 13			Age 17#		
	1969-70	1973-74	1978-79	1969-70	1973-74	1978-79	1969-70	1973-74	1978-79
Nation	2,142	2,614	2,510	2,226	2,556	2,762	1,994	2,301	2,702
Region									
Northeast	546	621	580	584	626	675	557	549	642
Southeast	435	634	625	523	625	657	406	555	683
Central	509	751	665	533	718	752	490	673	725
West	601	608	639	585	587	681	540	524	649
Sex									
Male	1,073	1,313	1,255	1,081	1,297	1,370	970	1,127	1,312
Female	1,068	1,301	1,255	1,129	1,259	1,395	1,024	1,174	1,386
Race									
White	1,606	2,051	1,849	1,666	2,03	2,053	1,649	1,907	2,134
Black	341	417	484	386	393	507	230	295	392
Other*	195	146	177	174	132	202	115	99	176
Parental education									
Not graduated high school	239	274	191	335	389	326	417	447	386
Graduated high school	502	585	589	712	766	895	657	745	895
Post high school	781	772	880	957	1,002	1,200	868	1,021	1,334
Unknown*	620	983	850	222	399	341	32	88	87
Type of community									
Extreme rural	216	207	252	225	245	271	203	236	246
Disadvantaged urban	211	314	250	230	299	281	195	289	306
Advantaged urban	209	202	253	210	314	277	197	210	265
Other communities	1,506	1,891	1,755	1,561	1,698	1,933	1,399	1,566	1,885
Size of community									
Big cities	593	670	723	597	717	775	536	546	718
Fringes around big cities	349	435	448	352	399	588	315	340	571
Medium cities	293	365	237	346	392	287	320	349	287
Small places	905	1,144	1,097	929	1,048	1,115	823	1,066	1,123
Grade									
3, 7, 10	494	571	623	546	707	683	256	299	360
4, 8, 11	1,559	1,976	1,834	1,585	1,764	2,031	1,379	1,705	2,014
12							219	251	285
Other*	89	67	53	95	85	48	140	46	43

+Data may not total due to rounding error.
*Data are not reported for these groups.
#17-year-olds enrolled in school.

TABLE D-2. Estimated Current Population Proportions of National Assessment Reporting Groups for In-School Students

Reporting Groups	9	13	17
Sex			
Male	.50	.50	.48
Female	.50	.50	.52
Race			
White	.79	.80	.83
Black	.14	.13	.12
Other	.07	.07	.05
Region			
Northeast	.25	.25	.25
Southeast	.22	.23	.20
Central	.27	.27	.29
West	.26	.25	.26
Parental education			
Not graduated high school	.09	.13	.15
Graduated high school	.24	.32	.32
Post high school	.33	.42	.48
Unknown	.34	.13	.05
Type of community			
Rural	.08	.10	.08
Disadvantaged urban	.07	.07	.09
Advantaged urban	.11	.11	.11
Other communities	.74	.72	.72
Size of community			
Big cities	.20	.21	.19
Fringes around big cities	.22	.22	.26
Medium cities	.12	.11	.11
Small places	.46	.46	.44
Grade in school			
<3, <7, <10	<.01	.02	.02
3, 7, 10	.23	.25	.13
4, 8, 11	.75	.72	.75
<4, <8, 12	<.01	<.01	.10
Other	<.01	<.01	<.01

APPENDIX E

ADJUSTMENT OF RESPONDENT WEIGHTS BY SMOOTHING TO REDUCE RANDOM VARIABILITY OF ESTIMATED POPULATION PROPORTIONS

Background

As noted elsewhere, a weight is assigned to every individual who responds to an exercise administered in an assessment. The weight is the reciprocal of the probability of selection of the individual with adjustment for nonresponse, and the probabilities of selection are based on the estimated number of people in the target age populations. Therefore, the weight for an individual estimates the number of people that the individual represents in the age population. The sum of the weights of all individuals at an age level who responded to an exercise is an estimate of the total number of people in that age population in the year that the exercise was assessed. Similarly, the sum of weights for all individuals who took the exercise and who also are members of some demographic category (such as blacks) gives an estimate of the number of people in the age population, for the year, who are also members of the category. The ratio of the two totals estimates the proportional representation of the demographic category in the age population for the given year.

Separate estimates of the proportional representation of the various demographic subgroups are provided by each booklet administered to a particular age group in a given year. Due to random sampling variability, the estimates of population proportions for a given year based on single booklets will vary. There is also random sampling variation in estimates of population proportions from year to year in addition to whatever trends in population proportions over time that might exist.

It is desirable to reduce the random variability of population proportions as much as possible since this variability has an effect on performance estimates. For example, the percentage of acceptable responses for an age group is a function of the relative proportions of high-performing and low-performing groups. If the relative proportions of these groups are very different in different assessments due to sampling variability, then a portion of the change in percentage of acceptable responses for an age group could be attributable to yearly sampling differences in the relative proportions of high- and low-achieving groups.

In addition to reporting performance estimates for an age group as a whole, National Assessment also reports performance for various subpopulations, such as whites or blacks. Because variability of subgroups

within these subpopulations (such as males and females within the white subpopulation) influence the performance estimates for the subpopulations, it is desirable that fluctuations of proportions of all subgroups of each subpopulation be reduced as much as possible.

For each age and year, each of the various booklets administered provides estimates of a given population proportion. Since these estimates are subject to booklet-to-booklet variability, a better estimate of the population proportion, which will have reduced variability, is obtained by combining the information from all booklets. However, these proportions vary from year to year due to random sampling variability or systematic differences in sampling procedures. An even better estimate of population proportions for any single year can be obtained by smoothing the proportions over several assessment years. The word "smoothing" is used here in the sense of fitting a smooth curve to a sequence of numbers by robust/resistant procedures (Tukey, 1977). Smoothing estimates of population proportions reduces a large portion of the sampling variability while preserving, as far as possible, actual trends occurring in the age population.

After the population proportions have been smoothed, adjusted weights are derived for the assessed individuals so that the population proportions computed using the adjusted weights are equal to the smoothed proportions. The adjusted weights are then used for all analyses.

Smoothing Procedures Used by National Assessment

The most direct way to smooth proportions is first to classify people into mutually exclusive multiway cells on the basis of their membership in categories of various important variables and then to smooth the proportions within each of the resulting multiway cells across years. Unfortunately, this procedure tends to produce a large number of cells with few people and, consequently, quite unstable estimates of smoothed proportions.

To circumvent this difficulty, National Assessment has utilized various smoothing procedures. These procedures, which are all basically weighting-class adjustments applied independently to each age, are designed to control, to varying degrees, fluctuations in certain key subgroups while avoiding, as much as possible, instabilities due to small cells.

The procedure used in 1978-79 has the following characteristics:

1. It produces a single adjusted weight for each individual.

2. It affords good control on the distribution of proportions of certain key variables.

3. It produces the greater stability of performance estimates than other procedures that have been used.

4. It is relatively easy to implement.

Even though adjusted weights using this procedure differ slightly from the corresponding adjusted weights from the other procedures that have previously been employed, National Assessment intends to use weights obtained using the 1978-79 procedure for all future analyses of data assessed in earlier years. This is simply because we believe weights obtained through this procedure to be the best available.

The Current Smoothing Procedure

The first step in the 1978-79 smoothing procedure involved the partitioning of the population of age class eligibles into the six smoothing cells given in Table E-1. The same cells were used for all ages.

TABLE E-1. Smoothing Cells Used for the 1978-79 Smoothing Procedure

Cell	Race	Region	Community Size (CS)
1	White	All	Big city + fringe (BC+FR)
2	White	All	Medium city (MC)
3	White	All	Small places (SP)
4	Black	SE	All
5	Black	Not SE	All
6	Other	All	All

Then, for each age and every year, the proportion of the population in each of the cells was estimated. For a given age and year, the proportion of the population in a particular cell was computed as the sum of weights of all respondents assessed in the given year who were of the specified age and who belonged in the cell, divided by the total of the weight of all respondents of the given age assessed in that year.

Each of the six cells was comprised of a sequence of estimated population proportions corresponding to the various years of assessment. Each such sequence of proportions was then smoothed by fitting robust/resistant lines. Using data from the U.S. Census and Current Population Surveys, trends in enrollment by age and race and by age and region were obtained. The data from these surveys were adjusted to correspond with NAEP definitions as much as possible. The resistant lines within the smoothing cells were constrained to satisfy the trend from the U.S. Census and Current Population Surveys data.

The final step in the smoothing procedure was to adjust the respondents' weights to be consistent with the smoothed proportions. Since each respondent

takes only one booklet, the weight adjustments were done independently for each booklet. For a given age, year and booklet, population proportions using the original weights were obtained for each of the smoothing cells. Then the weights of all respondents of a given cell were multiplied by the ratio of the smoothed cell proportion to the proportion using the original weights. This produced the adjusted weights that were used in all analyses.

Adjustment of Weights by Users

The smoothed population proportions for 9-, 13- and 17-year-olds (in-school only) are given in Tables E-2, E-3 and E-4, respectively. The columns of each table represent the smoothing cells while the rows represent the assessment year. For example, the smoothed population proportion of 9-year-olds in smoothing cell 2 (whites in medium cities) for 1972-73 is .1158.

To adjust respondent weights to be consistent with the smoothed proportions, the following procedures were followed:

1. For each booklet, respondents were classified according to smoothing cell, and the raw population proportions for each cell were obtained. For example, the raw proportions for a booklet given to 9-year-olds in smoothing cell 4 was the total of the weights of all 9-year-olds

---

TABLE E-2. Smoothed Frequencies From 10-Year Smooth by Smoothing Cell and Year for 9-Year-Olds

	Cell					
	1	2	3	4		6
Race	White	White	White	Black	[illegible] SE	Other
Region	All	All	All	SE		All
Size of Community	BC+FR	MC	SP	All	All	All
Year						
1969-70	0.3293	0.1258	0.3546	0.0535	0.0745	0.0624
1970-71	0.3251	0.1224	0.3601	0.0547	0.0746	0.0631
1971-72	0.3210	0.1191	0.3656	0.0558	0.0747	0.0639
1972-73	0.3168	0.1158	0.3711	0.0570	0.0748	0.0646
1973-74	0.3126	0.1124	0.3766	0.0581	0.0748	0.0654
1974-75	0.3085	0.1091	0.3822	0.0592	0.0749	0.0661
1975-76	0.3043	0.1058	0.3877	0.0604	0.0750	0.0668
1976-77	0.3001	0.1024	0.3932	0.0615	0.0751	0.0676
1977-78	0.2959	0.0991	0.3987	0.0627	0.0752	0.0683
1978-79	0.2918	0.0958	0.4042	0.0638	0.0753	0.0691

TABLE E-3. Smoothed Frequencies From 10-Year Smooth by Smoothing Cell and Year for 13-Year-Olds

	Cell					
	1	2	3	4	5	6
Race	White	White	White	Black	Black	Other
Region	All	All	All	SE	Not SE	All
Size of Community	BC+FR	MC	SP	All	All	All
Year						
1969-70	0.3200	0.1309	0.3703	0.0513	0.0673	0.0602
1970-71	0.3201	0.1256	0.3731	0.0521	0.0681	0.0610
1971-72	0.3202	0.1202	0.3760	0.0528	0.0690	0.0618
1972-73	0.3203	0.1149	0.3788	0.0536	0.0698	0.0626
1973-74	0.3203	0.1096	0.3816	0.0544	0.0714	0.0642
1974-75	0.3205	0.1043	0.3845	0.0552	0.0714	0.0642
1975-76	0.3206	0.0989	0.3873	0.0560	0.0722	0.0650
1976-77	0.3207	0.0936	0.3901	0.0567	0.0731	0.0658
1977-78	0.3207	0.0883	0.3929	0.0575	0.0739	0.0666
1978-79	0.3208	0.0830	0.3958	0.0583	0.0747	0.0674

TABLE E-4. Smoothed Frequencies From 10-Year Smooth by Smoothing Cell and Year for 17-Year-Olds

	Cell					
	1	2	3	4	5	6
Race	White	White	White	Black	Black	Other
Region	All	All	All	SE	Not SE	All
Size of Community	BC+FR	MC	SP	All	All	All
Year						
1969-70	0.3405	0.1447	0.3686	0.0415	0.0581	0.0466
1970-71	0.3419	0.1386	0.3704	0.0427	0.0591	0.0472
1971-72	0.3432	0.1326	0.3722	0.0440	0.0600	0.0478
1972-73	0.3446	0.1265	0.3740	0.0452	0.0612	0.0484
1973-74	0.3460	0.1204	0.3759	0.0465	0.0622	0.0491
1974-75	0.3474	0.1143	0.3777	0.0477	0.0633	0.0497
1975-76	0.3487	0.1082	0.3795	0.0490	0.0643	0.0503
1976-77	0.3501	0.1021	0.3813	0.0502	0.0653	0.0509
1977-78	0.3515	0.0961	0.3831	0.0515	0.0664	0.0515
1978-79	0.3529	0.0900	0.3849	0.0527	0.0674	0.0522

in the booklet who were black and in the Southeastern region, divided by the total of the weights of all respondents to the booklet.

2. For each booklet and smoothing cell, a weight adjustment factor as the ratio of the smoothed population proportion (for the appropriate age, year and smoothing cell) over the raw population proportion was obtained.

3. The adjusted weights for an individual were the product of that individual's original weight and the appropriate adjustment factor.

Changes in Smoothed Proportions as New Assessments Are Completed

Every time an assessment is completed, a new time point is added to each of the sequences of population proportions within the smoothing cells. This means that, even though robust/resistant procedures are used, the addition of a new point may somewhat change the values of smoothed proportions for prior years. Additionally, any changes in methodology will impact the estimates.

This means that the smoothed proportions, with the addition of the next assessment data, are apt to differ somewhat from the corresponding smoothed proportions without the new data. National Assessment has adopted the philosophy that the smoothed proportions, based on all currently available data using the best available algorithm, are the best available. Therefore, all subsequent analyses, for any year, will be done using this best available information, even though this may produce estimates that slightly differ from prior values.

GLOSSARY OF NATIONAL ASSESSMENT TERMS

Acceptable response. Any response to an exercise that demonstrates achievement of the objective measured by that exercise.

Administration time. The total time allowed on the paced audio tape for an exercise. (Includes the time allowed for the stimulus and the response.)

Administration timetable. Time periods during the school year when the various age groups are assessed. The time periods are:

October-December	13-year-olds
January-February	9-year-olds
March-May	17-year-olds

Age group or age level. Three age groups have been sampled in all three writing assessments: 9-year-olds, 13-year-olds and 17-year-olds attending school. Birth date ranges for each age group in each of the three assessments are as follows:

Assessment	Age 9	Age 13	Age 17
1969-70	1960	1956	10/51-9/52
1973-74	1964	1960	10/56-9/57
1978-79	1969	1965	10/61-9/62

Assessment. The documentation of the progress in knowledge, skills and attitudes of American youth. Measures are taken at periodic intervals for each learning area, with the goal of determining trends and reporting the findings to the public and to the education community.

Assessment administrator. Individual employed to administer the assessment in participating schools.

Background questions. Questions about respondents' instructional experiences with writing and frequency of their writing were included in writing item booklets. Standard background questions asked in every learning area are found on the back pages of the item booklets and include such things as level of parental education and reading materials in the home. Background questions used in the 1978-79 writing assessment appear in Appendix B.

Booklet. Items (exercises) are presented to respondents in booklets. Booklets are designed to be scored by optical scanning machines. Each booklet contains (1) instructions for answering items and sample items, (2) assessment items and (3) background questions. Each booklet contains approximately 30-35 minutes of assessment items and 10-15 minutes of

introductory material and background questions.

Category (scoring). A classification of a response to an open-ended exercise. See scoring guide.

Category within a variable. A subclassification within a variable. For example, male and female are categories of the variable sex. Also see reporting groups.

Difficulty level. The percentage of acceptable responses to an exercise.

Exercise. A task designed to measure an objective. Because NAEP does not administer "tests," but instead describes education achievement over time, the term "exercise" is often used instead of the term "item" or "test item." The terms "item" and "exercise" are used synonymously in this report.

Exercise booklet. See booklet.

Exercise part. See item part.

Exercise pool. The entire set of exercises prepared for a learning area. This set includes recycled exercises, exercises developed for previous assessments but not used due to exercise booklet or budgetary constraints, and newly developed exercises.

Field test. A pretest of exercises to obtain information regarding clarity, difficulty levels, timing, feasibility and special administrative problems needed for revision and selection of exercises to be used in the assessment.

Grade in school. Results are reported for 9-year-olds in the 3rd and 4th grades; 13-year-olds in the 7th and 8th grades; and 17-year-olds in the 10th, 11th and 12th grades.

Group administration. Booklets were administered to groups of 10 to 25 students in 1978-79. In 1969-70 and 1973-74, group size was 12 respondents. A paced audio tape was used to provide uniform instructions and oral presentation of exercises.

Hand scoring (scoring). The coding of responses in a format compatible with the optical scanning equipment being used. Multiple-choice exercises can be directly machine scored; however, responses to open-ended exercises must be coded in scoring ovals so that they can then be machine scored. See scoring guide.

ID number. An identification number referring to the unique number assigned to each respondent. This number is assigned to preserve the anonymity of each respondent. NAEP does not keep records of the names of any individuals.

Item. See exercise.

Item booklet. See booklet.

Item part. Each part of an item that asks a separate question. Parts may all pertain to one stimulus, such as a graph or a table, or may concern the same topic.

Jackknife. The name of the algorithm used by NAEP to estimate standard errors of percentages and other statistics.

Learning area. One of the areas assessed by National Assessment: reading/literature, writing, mathematics, science, citizenship/social studies, art, music, career development. Also called "subject area."

Level of parental education. These levels are described in Appendix A.

Modal grade. The grade in which the majority of each in-school age group is enrolled. For 9-year-olds, the modal grade is the 4th grade; for 13-year-olds, the 8th grade and for 17-year-olds, the 11th grade.

Objective. A desirable education goal agreed upon by scholars in the field, educators and concerned lay persons, and established through the consensus approach.

Objectives redevelopment. After the initial assessment of a learning area, one of the first steps in preparing for reassessment is a review of the learning-area objectives. This is carried out by scholars in the field, educators and concerned lay persons. These reviews may result in revision, modification or total rewriting of the learning-area objectives to reflect current curricular goals and emphases; they may also result in the endorsement of the objectives from the previous assessment as adequate for the next assessment.

Open-ended exercise. A nonmultiple-choice exercise that requires some type of written or oral response.

Paced audio tape. A tape recording that accompanies each booklet to assure uniformity in administration. Instructions and exercises are played back from the tape recording so that reading difficulties will not interfere with an individual's ability to respond. Response time is included on the tape.

Primary sampling unit (PSU). First-stage sampling units, typically a county or a group of contiguous counties.

Principal's questionnaire. A data collection form given to school officials. The officials respond to questions concerning enrollments, size of the community, occupational composition of the community, and so forth. Samples of these questionnaires are found in Appendix B. See also supplementary principal's questionnaire.

PSU. See primary sampling unit.

Public-use data tapes. Computer tapes containing respondent-level exercise and background/demographic data and machine-readable documentation. These tapes are available for use by external researchers wishing to do secondary analyses of National Assessment data.

Racial/ethnic category. For the writing assesments, results are reported for whites and blacks.

Receipt control. Procedures implemented by scoring staff to check in and screen materials from the field. Information gained from receipt control procedures is relayed to assessment administrative staff so that any errors may be corrected.

Recycled exercises. The set of exercises that is kept secure from one assessment to the next that will be used to measure change (growth, stability or decline) in performance for the learning area.

Region. One of four geographical regions used in gathering and reporting data: Northeast, Southeast, Central and West. States included in each region are shown in Appendix A.

Released exercise. An exercise for which results and exercise text have been reported to the public.

Released exercise set. A set of released exercises, including documentation and scoring guides, that can be purchased from National Assessment.

Reporting groups. Categories of variables for which National Assessment data are reported. Variable categories are defined in Appendix A.

Rescore. If an open-ended exercise was scored under different conditions than presently held or if passage of time might affect scoring, responses from a previous assessment may be rescored at the same time that responses from a later assessment are scored. Responses from an earlier assessment also may be held and not scored so that responses from a later assessment can be scored at the same time.

Respondent. A person who responds to the exercises in an assessment booklet.

Response options. Different alternatives to a multiple-choice question that can be selected by the respondent.

Review conference. A conference held to review the objectives of a learning area to assure their acceptance as measures of the objectives by scholars, educators and lay persons or to review exercises for racial, ethnic, social or regional bias.

Sample. National Assessment does not assess an entire age population but rather selects a representative sample from the age group to answer assessment items. (See Chapter 4 for a description of National Assessment sampling procedures.)

Scoring guide. A guide for hand scoring an open-ended exercise that specifies descriptive or diagnostic categories by giving definitions and sample responses.

Scoring ovals. Scannable ovals printed beside multiple-choice options and printed at the bottom of the page for open-ended exercises (to be used in handscoring). When ovals are marked, they can be scored by machine and responses recorded by computer.

Sex. Results are reported for males and females.

Size of community. Results are reported for four size-of-community categories: big cities, fringes around big cities, medium cities and small places. These categories are defined in Appendix A.

SMSA. Standard Metropolitan Statistical Area. SMSAs are economic and social units defined by the U.S. Bureau of the Census.

Standard error. A measure of sampling variability for a statistic. Because of NAEP's complex sample design, standard errors are estimated by jackknifing first-stage sample estimates.

Stem. The portion of an exercise that states the problem or asks the question.

Stimulus. See stem.

Subject area. See learning area.

Subpopulation or subgroup. Groups within the national population, such as males and females, for which data are reported.

Supplementary principal's questionnaire. A data collection form given to school officials. On this form, officials are asked to respond to questions concerning course offerings, materials and staffing specific to the learning area being assessed. A sample of this questionnaire is found in Appendix B. See also principal's questionnaire.

Tapescript. A script prepared for the announcer to use in producing the paced tape. It indicates exactly what is to be read or not read and indicates the amount of response time allowed for each exercise. See paced audio tape.

Timing. Most NAEP exercises are administered with a paced audio tape to standardize data collection conditions. The tape includes the amount of time students are allowed to respond to each exercise.

Type of community. Results are reported for three type-of-community categories: disadvantaged urban, advantaged urban and rural. Definitions of these categories are found in Appendix A.

User tape. See public-use data tape.

Variable. A classification of respondents. Standard reporting variables are: region, sex, race, level of parental education, size of community, type of community and grade in school.

Weight. A multiplicative factor equal to the reciprocal of the probability of a respondent being selected for assessment with adjustment for nonresponse -- an estimate of the number of persons in the population represented by a respondent in the sample. Theoretically, the sum of weights for all respondents at an age level is equal to the number of persons in the country at that age level.

BIBLIOGRAPHY

Brown, Rexford G. "Evaluations of Writing: Some Theoretical Considerations and Practical Suggestions." New York: Ford Foundation (forthcoming).

Chromy, J. and D. Horvitz. "Structure of Sampling and Weighting," 1969-70 Science: National Results and Illustrations of Group Comparisons, Report no. 1, 1969-70 Assessment. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1970. ERIC no. ED 055 786. ISBN 0-89398-276-8.

Final Report on National Assessment of Educational Progress Sampling and Weighting Activities for Assessment Year 10. Research Triangle Park, N.C.: Research Triangle Institute, 1980.

Final Report on Year 09 Field Operations and Data Collection Activities, National Assessment of Educational Progress. Research Triangle Park, N.C.: Research Triangle Institute, 1978.

Final Report on Year 09 In-School Field Operations and Data Collection Activities, National Asessment of Educational Progress. Research Triangle Park, N.C.: Research Triangle Institute, 1978.

Folsom, Ralph E. National Assessment Approach to Sampling Error Estimation, Sampling Error Monograph 25U-796-5. Research Triangle Park, N.C.: Research Triangle Institute, 1977.

Kalsbeek, W.D. et al. No Show Analysis, final report. Raleigh, N.C.: Research Triangle Institute, 1975.

Miller, R.G. Jr. "A Trustworthy Jackknife," Annals of Mathematical Statistics. no. 35, 1964.

Miller, R.G. Jr. "Jackknifing Variances," Annals of Mathematical Statistics, no. 39, 1968.

Miller, R.G. Jr. "The Jackknife -- A Review," Biometrika, no. 61, 1974.

Moore, R. et al. The National Assessment Approach to Sampling. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1974. ERIC no. ED 099 416.

Mosteller, F. and J.W. Tukey. Data Analysis and Regression. Reading, Mass.: Addison-Wesley, 1977.

Mosteller, F. and J.W. Tukey, "Data Analysis Including Statistics," in Handbook of Social Psychology, 2nd ed., edited by E. Aronson and G. Lindzey. Reading, Mass.: Addison-Wesley, 1968.

Mullis, Ina V.S. "Using the Primary Trait System for Evaluating Writing," no. 10-W-51. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980. ISBN 0-89398-396-9.

Mullis, Ina V.S. and John C. Mellon. "Guidelines for Three Ways of Evaluating Writing: Syntax, Cohesion and Mechanics," no. 10-W-50. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980. ISBN 0-89398-395-0.

Rogers, W.T. et al. "Assessment of Nonresponse Bias in Sample Surveys: An Example From National Assessment," Journal of Educational Measurement, vol. 14, no. 4, 1977.

The Third Assessment of Writing: 1978-79 Released Exercise Set, no. 10-W-25, 1978-79 Assessment. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1981. ISBN 0-89398-394-2.

Tukey, John W. Exploratory Data Analysis. Reading, Mass.: Addison-Wesley, 1977.

Writing Achievement, 1969-79: Results From the Third National Writing Assessment, Volume I -- 17-Year-Olds, Report no. 10-W-01, 1969-70, 1973-74 and 1978-79 Assessments. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980. ISBN 0-89398-400-0.

Writing Achievement, 1969-79: Results From the Third National Writing Assessment, Volume II -- 13-Year-Olds, Report no. 10-W-02, 1969-70, 1973-74 and 1978-79 Assessments. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980. ISBN 0-89398-401-9.

Writing Achievement, 1969-79: Results From the Third National Writing Assessment, Volume III -- 9-Year-Olds, Report no. 10-W-03, 1969-70, 1973-74 and 1978-79 Assessments. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1980. ISBN 0-89398-402-7.

Writing Mechanics, 1969-74, Report no. 05-W-01, 1969-70 and 1973-74 Assessments. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1975. ERIC no. ED 113 736. ISBN 0-89398-383-7.

Writing Objectives, Second Assessment, 1973-74 Assessment. Denver, Colo.: National Assessment of Educational Progress, Education Commission of the States, 1972. ERIC no. ED 072 460. ISBN 0-89398-381-0.

(Continued From Inside Front Cover)

In addition to the above reports, National Assessment has produced reports in the areas of social studies, citizenship, music, art mathematics, science and career and occupational development. A complete publications list and ordering information are available from the address below.

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
Suite 700, 1860 Lincoln Street
Denver, Colorado 80295